目　录

第一章　引　论

本章首先引进各种广义逆矩阵的定义和一些必要的记号，在此基础上，概述广义逆矩阵的理论和应用的重要历史发展和现状.

§1.1　广义逆矩阵的定义

广义逆矩阵的概念渊源于线性方程组的求解问题.

设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，$\boldsymbol{x}$ 为 $n\times 1$ 未知向量，$\boldsymbol{b}$ 为 $m\times 1$ 常数向量. 众所周知，当 $m=n$ 且 $\boldsymbol{A}$ 为可逆阵时，线性方程组

$$\boldsymbol{Ax}=\boldsymbol{b} \tag{1.1.1}$$

有唯一解，并且这个唯一解可以表示为

$$\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}. \tag{1.1.2}$$

然而，在许多情况下，方程组（1.1.1）的系数矩阵 $\boldsymbol{A}$ 可能是奇异阵，或者根本不是方阵，但它的解却是存在的. 在这种情况下，我们如何用类似于（1.1.2）的形式来表示（1.1.1）的解呢？更进一步，若（1.1.1）是不相容的，即该方程组无解，这时我们退而求其次，改为求它的最小二乘解，也就是求极小化

$$\|\boldsymbol{b}-\boldsymbol{Ax}\|^2=(\boldsymbol{b}-\boldsymbol{Ax})^*(\boldsymbol{b}-\boldsymbol{Ax})$$

的向量 $\boldsymbol{x}$. 以上这些问题，以及在数理统计、数学规划、计量经济、数值分析、控制论和网络等领域的许多其他问题都需要推广通常逆矩阵的概念，这就产生了所谓的广义逆矩阵.

对任意一个 $m\times n$ 矩阵 $\boldsymbol{A}$，Penrose（1955）用下面的四

个方程定义 A 的广义逆：

$$
\begin{aligned}
&(1) \qquad & AXA = A, \\
&(2) \qquad & XAX = X, \\
&(3) \qquad & (AX)^* = AX, \\
&(4) \qquad & (XA)^* = XA,
\end{aligned}
\qquad (1.1.3)
$$

其中 A^* 表示 A 的转置共轭阵. Penrose (1955) 证明了，满足上面四个方程的矩阵 X 是唯一的. 然而，在更早些时候，Moore (1920, 1935) 用另一种等价形式定义了这种广义逆 (见 § 1.2). 后来人们就把(1.1.3)定义的广义逆称为 Moore-Penrose 逆，简记为 A^+，而把 (1.1.3) 的四个矩阵方程统称为 Penrose 方程.

事实上，若仅仅为了表示 (1.1.1) 的解，我们只需要满足 (1.1.3) 中的方程 (1) 的矩阵 X，因此我们可以定义一种只满足方程 (1) 的广义逆. 更一般地，为了不同的目的，人们定义了满足 Penrose 方程中任意若干个方程的广义逆.

记 $C^{m\times n}$和 $R^{m\times n}$分别为 $m\times n$ 复矩阵和实矩阵的全体.

定义 1.1.1 对任意 $A\in C^{m\times n}$，用 $A\{i, j, \cdots, l\}$ 表示 Penrose 方程 (1.1.3) 中方程 (i), (j), …, (l) 的解集. 若 $X\in A\{i, j, \cdots, l\}$，则称 X 为 A 的一个$\{i,j,\cdots,l\}$-逆，记之为 $A^{(i,j,\cdots,l)}$.

下面是几个简单例子.

例 1.1.1 设 A 为 $n\times n$ 方阵且秩 $r(A)=n$，这时 A 的逆矩阵 A^{-1}存在. 容易验证，它满足 Penrose 方程，因此 A^{-1}就是一个 $A^{(1,2,3,4)}$，即 A^+. 我们将证明，对任意的矩阵 A，A^+ 存在且唯一. 于是对可逆矩阵 A，$A^{-1}=A^+$.

例 1.1.2 设 a 为 $n\times 1$ 向量，则不难验证

$$x = \frac{a'}{\|a\|^2}$$

是 a^+，其中 $\|a\|^2=a^* a$.

例 1.1.3 设 $m\times n$ 矩阵 A 为行满秩阵，即 $r(A)=m$. 此时 AA^* 为 $m\times m$ 可逆阵. 记

$$X = A^*(AA^*)^{-1},$$

则有

$$AX = AA^*(AA^*)^{-1} = I_m, \tag{1.1.4}$$

其中 I_m 表示 m 阶单位阵. 我们称 X 为 A 的右逆阵. 利用(1.1.4)，很容易看到，X 满足 Penrose 方程的前三个方程. 于是，对任一行满秩矩阵 A，它的右逆阵 $X=A^*(AA^*)^{-1}$是一个 $A^{(1,2,3)}$.

例 1.1.4 设 $m\times n$ 矩阵 A 为列满秩阵，即 $r(A)=n$. 此时 A^*A 为 $n\times n$ 可逆阵. 记

$$X = (A^*A)^{-1}A^*,$$

则有

$$XA = (A^*A)^{-1}A^*A = I_n, \tag{1.1.5}$$

我们称 X 为 A 的左逆阵. 利用(1.1.5)容易验证，X 满足 Penrose 方程的 (1)、(2) 和 (4). 于是，对于任一列满秩矩阵 A，它的左逆阵 $X=(A^*A)^{-1}A^*$ 是一个 $A^{(1,2,4)}$.

为符号简单计，像大多数文献一样，我们把 $A^{(1)}$记为 A^-. 前面我们已经把 $A^{(1,2,3,4)}$记为 A^+. 对于任一矩阵 A，在其广义逆矩阵中，A^-和 A^+是最重要的两种广义逆，将在第三章和第四章分别详细讨论.

广义逆 $A^{(1,2)}$，又称为自反广义逆 (reflexive generalized inverse)，此因在 Penrose 方程 (1) 和 (2) 中，A 和 X 的地位完全对称，所以 $X\in A\{1, 2\}$ 与 $A\in X\{1, 2\}$ 必同时成

立.

广义逆 $A^{(1,3)}$，又称为最小二乘广义逆. 它在数值分析与数理统计中有许多特殊应用. 设 (1.1.1) 是不相容方程组，若 $\hat{x}$ 满足

$$\| b - A\hat{x} \| = \inf_x \| b - Ax \| ,$$

则称 $\hat{x}$ 为不相容方程组的最小二乘解. $A^{(1,3)}$ 的重要性在于 $x = A^{(1,3)}b$ 是一个最小二乘解，这就是"最小二乘广义逆"名称的由来.

广义逆 $A^{(1,4)}$，也称为最小范数广义逆. 设 (1.1.1) 是相容方程组，我们将证明(见 § 3.3)，对任意一个 A^-，$x=A^-b$ 都是解. 在这个解集中，$x=A^{(1,4)}b$ 的范数最小，即

$$\| A^{(1,4)}b \| = \inf_{A^-} \| A^- b \| .$$

以上这三种广义逆以及其他一些 $A\{i, j, \cdots, l\}$ 将在第五章讨论.

在文献中还存在另外一些广义逆矩阵，主要有群逆 (group inverse)、Drazin 逆和 Bott-Duffin 逆.

定义 1.1.2 设 $A \in C^{n\times n}$，若 n 阶方阵 $X \in A\{1, 2\}$ 且满足 $AX=XA$，则称 X 为 A 的群逆，记为 $A^{(g)}$.

这个名称是由 Erdelyi (1967) 引进的，其原因是 A 和 $A^{(g)}$ 的正数次幂 A^k 和 $(A^{(g)})^k$ 与 $AA^{(g)}$ 一起构成一个 Abel 群，这里 $AA^{(g)}$ 是这个群的单位元. 从定义 1.1.2 可以看出，群逆 $A^{(g)}$ 是一个特殊的 $A^{(1,2)}$，但是这种广义逆不是总存在.

为了引进下面的 Drazin 逆，我们需要方阵指数的概念. 设 A 为方阵，我们称满足

$$r(A^k) = r(A^{k+1})$$

的最小正整数 k 为 A 的指数 (index)，其中 $r(A)$ 表示矩阵

A 的秩.

定义 1.1.3 设 $A\in C^{n\times n}$，其指数为 k. 若方阵 X 满足

$$A^kXA = A^k,$$
$$XAX = X,$$
$$AX = XA,$$

则称 X 为 A 的 Drazin 逆.

这种广义逆是由 Drazin 于 1958 年引进的. 易见，若 A 的指数为 1，则 A 的 Drazin 逆就是群逆.

记 C^n 为 $n\times 1$ 的复向量的全体. 设 S 为 C^n 的一个子空间，$S^{\perp}$ 为 S 的正交补空间，用 P_S 表示向 S 的正交投影阵.

定义 1.1.4 设 $A\in C^{n\times n}$ 且 $AP_S+P_{S^{\perp}}$ 可逆，则称矩阵

$$P_S(AP_S + P_{S^{\perp}})^{-1}$$

为 A 关于 S 的 Bott-Duffin 逆，记为 $A_{(S)}^{(-1)}$.

这种广义逆是 Bott 和 Duffin (1953) 研究电网络理论时提出的. 自然，当 $AP_S+P_{S^{\perp}}$ 不可逆时，我们可以用

$$(AP_S + P_{S^{\perp}})\{i,j,\cdots,l\}.$$

由于上述几种广义逆的应用尚不广泛，因此本书不讨论它们，对其感兴趣的读者可阅读 Ben-Israel 等 (1974).

§1.2 历史概略

Moore, E. H 是公认的研究广义逆矩阵的第一人. 他在美国数学会 1920 年一个会议报告的摘要中，对任意矩阵定义了广义逆，当时他称之为 general reciprocal. 但有的学者认为，Moore 关于广义逆的研究可以追溯到 1906 年，Moore 关于广义逆的较详细结果发表在 Moore (1935) 的著名论文中. 于是，许多学者常把 1935 年作为广义逆研究的起点. 在这篇论

文中，对任意 $m\times n$ 矩阵 A，Moore 用下面两个矩阵方程

$$AX = P_A, XA = P_X \tag{1.2.1}$$

来定义广义逆 X，这里 P_A 表示向矩阵 A 的列空间上的正交投影阵. 事实上，(1.2.1) 等价于 Penrose 方程 (1.1.3). 但是，在此后的 20 余年中，人们对广义逆的研究并未给予应有的重视.

到了本世纪 50 年代，一些学者开始注意到广义逆矩阵的最小二乘性质. Bjerhammar (1951a, 1951b) 在不知道 Moore 结果的情况下，重新提出了广义逆矩阵的概念（他称之为 reciprocal matrix），并注意到了广义逆与线性方程组解的关系. Bott 和 Duffin (1953) 在研究电网络理论时，引进了一种后来被称为 Bott-Duffin 广义逆的逆矩阵. 当时他们称为约束逆 (constrained inverse). 但这时期的研究工作缺少一般性，零散而不系统.

在广义逆研究中，一个重要的里程碑是 Penrose (1955) 的著名论文. 在这篇文章中，Penrose 用 (1.1.3) 的四个方程再次定义了广义逆，并证明了 (1.1.3) 的解是唯一的. 他还建立了 (1.1.3) 的第一方程的解 A^- 与方程组 (1.1.1) 解的联系. Penrose 的这项工作在广义逆研究中起着十分重要的作用，它使广义逆这一概念获得再生. 从此之后，学者们对广义逆的研究倾注了前所未有的兴趣. 在此后短短 10 余年中，发表了数百篇关于广义逆的研究论文. 这包括 Greville (1957), Bjerhammar (1957, 1958), Ben-Israel 和 Charnes (1963), Chipman (1964), Scroggs 和 Odell (1966) 等. 在这期间, Erdelyi (1967) 引进了群逆，而 Drazin 于 1958 年引进了另一种广义逆，他称为 pseudo inverse，现在通称为 Drazin 逆.

在广义逆研究的这一高峰时期，统计学家的研究工作占了相当的分量. Rao (1955, 1962, 1966, 1967) 和 Mitra (1968a, 1968b) 研究了{1}- 逆结构表示和不唯一性，并把它们应用于统计参数估计理论，特别是线性模型和方差分析的估计与检验问题. 现在广义逆矩阵已经成为数理统计的许多分支不可缺少的有效工具，参见王松桂 (1987), Wang 和 Chow (1994).

1968 年 3 月，在美国 Texas 举行了广义逆矩阵的专题学术会议，并出版了会议文集，见 Boullion 和 Odell (1968). 后来，分别于 1973 年和 1976 年举行了关于这一课题的讨论班和区域性会议，并出版有文集，分别见 Nashed (1976) 和 Campbell (1982).

Ben-Israel (1966), Stewart (1969), Wedin (1973) 和何旭初 (1979) 研究了广义逆矩阵的扰动和连续性问题，他们建立了 A^+ 连续性的条件.

在本世纪 70 年代前后，一些关于广义逆矩阵及其应用的专著陆续问世，其中主要有 Rao 和 Mitra (1971), Pringle 和 Rayner (1971), Boullion 和 Odell (1971), Ben-Israel 和 Greville (1974). 这些著作广泛收集和系统总结了散见在各种刊物中关于广义逆的理论、方法和应用的许多重要结果，并在一定程度上规范了许多常用的术语和记号.

由于统计研究工作的需要，近年来许多统计学家对涉及广义逆的偏序 (partial ordering) 显示了很大兴趣. 我们用 $A \geqslant 0$ 表示 A 为半正定实对称阵，若 $A \geqslant 0$, $B \geqslant 0$ 且两者为同阶阵，我们用 $A \geqslant B$ 或 $B \leqslant A$ 表示 $A-B \geqslant 0$. 众所周知，若 A 和 B 皆可逆，且 $A \geqslant B$，则必有 $A^{-1} \leqslant B^{-1}$. Wu (1980) 把这个结果推广到各种广义逆，建立了诸如 $A^- \leqslant B^-$, $A^{(1,2)} \leqslant B^{(1,2)}$

成立的条件. Milliken 等（1977）证明了 $r(A)=r(B)$ 时，$A \geqslant B$ 的充要条件为 $A^+ \leqslant B^+$. Liski 和王松桂（1996）则建立了 $A^{(2)} \geqslant B^{(2)}$ 的若干充要条件. Marshall 和 Olkin（1979）、Baksalary 和 Puntanen（1991）、Wang 和 Shao（1992）、Wang 和 Liski（1996）、Liu 等（1995）把 Cauchy-Schwarz 不等式和 Kantorovich 不等式推广到含广义逆矩阵的情形.

我们知道，矩阵是现代自然科学、工程技术乃至社会科学许多领域的一个不可缺少的数学工具，因此广义逆矩阵的应用也相当广泛. 可以这样说，凡是用到矩阵的地方，都有可能用到广义逆矩阵. Nashed（1976）和 Campbell（1982）介绍或综述了广义逆在许多方面的应用，其中包括数理统计、数学规划、数值分析、控制论、博奕论和计量经济等，部分内容的详细讨论可以在 Rao 和 Mitra（1971）以及 Ben-Israle 和 Greville（1974）中找到.

第二章　矩阵论基础

本章从方便广义逆矩阵讨论的角度叙述矩阵论的一些重要结果，并引进必要的记号. 为了节省篇幅，有些事实的证明或被略去了或只给出证明的梗概.

§2.1　线性空间及其分解

我们用 $\boldsymbol{C}^n$ 和 $\boldsymbol{R}^n$ 分别表示全体 n 维复向量和 n 维实向量组成的线性空间，$\boldsymbol{C}^1$ 和 $\boldsymbol{R}^1$ 分别简记为 $\boldsymbol{C}$ 和 $\boldsymbol{R}$.

设 S_1 和 S_2 为 $\boldsymbol{C}^n$ 的两个子空间. 则它们的和定义为

$$S = S_1 + S_2 = \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in S_1, \boldsymbol{y} \in S_2\}. \quad (2.1.1)$$

S 也是 $\boldsymbol{C}^n$ 的一个子空间，而且是包含 S_1 和 S_2 的最小子空间. 如果它们的公共部分

$$S_1 \cap S_2 = \{\boldsymbol{x} : \boldsymbol{x} \in S_1, \boldsymbol{x} \in S_2\} = \{\boldsymbol{0}\},$$

则称 S 为 S_1 与 S_2 的直接和，记为

$$S = S_1 \oplus S_2. \quad (2.1.2)$$

如果

$$\boldsymbol{C}^n = S_1 \oplus S_2, \quad (2.1.3)$$

则称 S_1 和 S_2 是互补的，并称 S_1 为 S_2 的补空间，反过来，称 S_2 是 S_1 的补空间. (2.1.2) 和 (2.1.3) 分别称为 S 和 $\boldsymbol{C}^n$ 的直接和分解. 直接和分解 (2.1.2) 的一个重要性质是：若 $\boldsymbol{u} \in S$，则存在唯一的分解

$$\boldsymbol{u} = \boldsymbol{x} + \boldsymbol{y},$$

其中 $\boldsymbol{x} \in S_1$，$\boldsymbol{y} \in S_2$.

直接和分解可以推广到多个子空间的情形

$$S = S_1 \oplus S_2 \oplus \cdots \oplus S_k.$$

此时，对任给的 $\boldsymbol{u} \in S$，存在唯一的分解

$$\boldsymbol{u} = \sum_{i=1}^{k} \boldsymbol{x}_i, \quad \boldsymbol{x}_i \in S_i, \quad i = 1, \cdots, k.$$

用 $\dim S$ 表示空间 S 的维数. 下面的定理建立了直接和分解式（2.1.2）和（2.1.3）中各空间维数之间的关系.

定理 2.1.1 对直接和分解式（2.1.2）和（2.1.3），分别有

$$\dim S = \dim S_1 + \dim S_2, \tag{2.1.4}$$

$$\dim S_1 + \dim S_2 = n. \tag{2.1.5}$$

这个事实的证明是容易的.

对一个 $m \times n$ 矩阵 $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$，$\boldsymbol{a}_i \in \boldsymbol{C}^m$，$i = 1, \cdots, n$，与其相联系的有如下两个重要子空间：

$$\mathscr{M}(\boldsymbol{A}) = \{\boldsymbol{x} \in \boldsymbol{C}^m : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{t}, \text{对任意 } \boldsymbol{t} \in \boldsymbol{C}^n\},$$

$$\mathscr{N}(\boldsymbol{A}) = \{\boldsymbol{x} \in \boldsymbol{C}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\},$$

易见，$\mathscr{M}(\boldsymbol{A})$是 $\boldsymbol{A}$ 的列向量张成的子空间，称为 $\boldsymbol{A}$ 的列空间或值域(range)，而称 $\mathscr{N}(\boldsymbol{A})$为 $\boldsymbol{A}$ 的零空间(null space).

从定义知，$\dim \mathscr{M}(\boldsymbol{A}) = r(\boldsymbol{A})$，这里 $r(\boldsymbol{A})$表示矩阵 $\boldsymbol{A}$ 的秩. 但是 $\mathscr{N}(\boldsymbol{A})$是齐次线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的解空间，其维数为 $n - r(\boldsymbol{A})$，于是我们有

定理 2.1.2 对任意 $\boldsymbol{A} \in \boldsymbol{C}^{m \times n}$，

$$\dim \mathscr{M}(\boldsymbol{A}) = r(\boldsymbol{A}),$$

$$\dim \mathscr{N}(\boldsymbol{A}) = n - r(\boldsymbol{A}).$$

设 $\boldsymbol{x}$，$\boldsymbol{y} \in \boldsymbol{C}^n$，用记号 $(\boldsymbol{x}, \boldsymbol{y})$ 表示向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的一个复值函数，且具有下列性质：

(a) $(x, y)=\overline{(y, x)}$，这里 $\bar{x}$ 表示 x 的共轭复数；

(b) $(\alpha_1 x_1+\alpha_2 x_2, y)=\alpha_1(x_1, y)+\alpha_2(x_2, y)$，其中 $x_1, x_2, y\in C^n$，$\alpha_1, \alpha_2\in C$；

(c) 对一切 $x\in C^n$，$(x, x)\geqslant 0$，且等号成立当且仅当 $x=0$.

我们称函数 (x, y) 为向量 x 和 y 的内积，并称 $(x, x)^{1/2}$ 为向量 x 的长度或范数，记作 $\|x\|$. 定义了内积的 C^n 和 R^n 分别称为酉空间和欧氏空间.

在 C^n 中，每个内积都具有如下形式：

$$(x,y) = y^* M x,$$

这里 M 为 $n\times n$ 正定 Hermite 阵. 而对于 R^n，则

$$(x,y) = y' M x.$$

特别，当 $M=I$ 时的内积称为标准内积. 以后如无特殊声明，我们所指的内积总是指标准内积. 在必要的时候，把一般内积 $(x,y)=y^* M x$ 记为 $(x,y)_M$，对应的范数记为 $\|x\|_M$，即

$$\|x\|_M = (x^* M x)^{1/2}.$$

定理 2.1.3 下列两个不等式成立：

(a) Cauchy-Schwarz 不等式

$$|(x,y)| \leqslant \|x\| \, \|y\|. \qquad (2.1.6)$$

(b) 三角不等式

$$\|x+y\| \leqslant \|x\| + \|y\|. \qquad (2.1.7)$$

注 关于 Cauchy-Schwarz 不等式，近年来人们做了许多形式的推广，详见王松桂，贾忠贞（1994, p. 216），Liu 和 Neudecker（1995），Wang 和 Liski（1996）.

若 $(x, y)=0$，则称 x 与 y 正交，记为 $x\perp y$. 若向量组 $a_1, \cdots, a_m$ 两两正交，即

$$(a_i, a_j) = 0, \quad i\neq j,$$

则称 $a_1, \cdots, a_m$ 为正交向量组. 若进一步假设 $\|a_i\|=1$, $i=1, \cdots, m$, 则称该向量组为标准正交组. 设 S 为酉空间 C^n 的一个子空间或集合, 且对任意 $y\in S$, 都有 $x\perp y$, 则称向量 x 与 S 正交, 记为 $x\perp S$.

设 S_1 和 S_2 为 C^n 的两个子空间或集合. 若对任意 $x\in S_1$, $y\in S_2$, 总有 $(x, y)=0$, 则称 S_1 与 S_2 正交, 记为 $S_1\perp S_2$.

容易证明, 不论 S 是集合还是子空间, 向量集合

$$\{x:\quad x\perp S, x\in C^n\}$$

总是一个子空间, 记为 $S^\perp$. 不难证明

$$S^\perp \perp S, \tag{2.1.8}$$

$$S \oplus S^\perp = C^n. \tag{2.1.9}$$

我们称 $S^\perp$ 为 S 的正交补空间, 并称 (2.1.9) 为 C^n 的直交分解, 这里需假定 S 为子空间. 读者可以验证, 若 $A\in C^{m\times n}$, 则

$$\mathcal{N}(A) = \mathcal{R}(A^*)^\perp, \tag{2.1.10}$$

$$\mathcal{R}(A^*) \oplus \mathcal{N}(A) = C^n. \tag{2.1.11}$$

§2.2 矩阵标准形

把矩阵化成不同意义下的标准形, 是矩阵论研究的一个重要问题, 它在矩阵理论及其应用中占有重要地位, 也是广义逆矩阵研究的一个很有用的工具.

设 $\mathscr{A}$ 表示一个集合, 若在其中定义了一种关系"～", 满足

(a) 自反性: 对任意 $A\in\mathscr{A}$, 有 $A\sim A$;

(b) 传递性: 若 $A\sim B$, $B\sim C$, 必有 $A\sim C$;

(c) 对称性：若 $A \sim B$，必有 $B \sim A$，

则称关系"～"是一种等价关系.

在矩阵论中，有三种重要的等价关系：相抵、相似和相合.

定义 2.2.1 设 $\boldsymbol{A}, \boldsymbol{B} \in C^{m \times n}$. 若存在 m 阶可逆阵 $\boldsymbol{P}$ 和 n 阶可逆阵 $\boldsymbol{Q}$，使得 $\boldsymbol{B}=\boldsymbol{PAQ}$，则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵.

容易验证，相抵关系是一种等价关系. 下面的定理给出了矩阵在相抵意义下的最简形式：相抵标准形.

定理 2.2.1（相抵标准形） 设 $\boldsymbol{A} \in C^{m \times n}$，且秩为 r，则存在可逆阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$，使得

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \qquad (2.2.1)$$

其中 $\boldsymbol{I}_r$ 为 r 阶单位阵.

本定理的证明见许以超(1966,p.199). 为了后面证明的需要,下面我们给出证明的要点,并引进必要的记号和术语.

将矩阵 $\boldsymbol{A}$ 经过适当行向量的如下三种初等变换：

(1) 交换两行的位置；

(2) 用一个非零数去乘某一行；

(3) 将某一行乘一个数加到另一行，

可把矩阵 $\boldsymbol{A}$ 化为行约简梯形，也就是满足下列四个条件的矩阵：

(1) 非零行的第一个非零元素为 1，称其为主导元素；

(2) 某列中若含有主导元素 1，则其余元素全为零；

(3) 所有元素皆为零的行排在矩阵的下部；

(4) 主导元素 1 排成阶梯形，即若第 j 行的主导元素 1 位于第 k_j 列，$j=1, \cdots, r$，则 $k_1<k_2<\cdots<k_r$.

如果对矩阵 A 的行约简梯形经过若干次列与列之间的

交换，则可化为如下分块矩阵的形式：

$$\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}. \tag{2.2.2}$$

因为对矩阵 $\boldsymbol{A}$ 施以一系列三种行初等变换相当于用一个可逆阵左乘矩阵 $\boldsymbol{A}$，而一系列的列互换相当于用置换矩阵右乘矩阵 $\boldsymbol{A}$，于是存在可逆阵 $\boldsymbol{B}$ 和置换阵 $\boldsymbol{C}_1$ 使得

$$\boldsymbol{BAC}_1 = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}. \tag{2.2.3}$$

令

$$\boldsymbol{C}_2 = \begin{pmatrix} \boldsymbol{I}_r & -\boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{I}_{n-r} \end{pmatrix},$$

则

$$\boldsymbol{BAC}_1\boldsymbol{C}_2 = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

取 $\boldsymbol{P}=\boldsymbol{B}$，$\boldsymbol{Q}=\boldsymbol{C}_1\boldsymbol{C}_2$，便得到（2.2.1）.

注 1　在相抵关系下，矩阵的秩是不变量.

注 2　行约简梯形也称为 Hermite 典则形. 对一个给定矩阵，它的行约简梯形不是唯一的. 然而文献中对行约简梯形还有另外一种定义，对这种定义，它是唯一的，请参阅倪国熙（1984）.

定义 2.2.2　对任意两个 n 阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，若存在可逆阵 $\boldsymbol{P}$，使得 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$，则称 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 相似.

容易验证，相似也是一种等价关系. 下面的定理给出了方阵在相似意义下的最简形式，即 Jordan 标准形.

定理 2.2.2（Jordan 标准形）　对任意一个 n 阶复方阵 $\boldsymbol{A}$，都存在可逆阵 $\boldsymbol{P}$，使得

$$P^{-1}AP = \begin{bmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_s}(\lambda_s) \end{bmatrix}, \quad (2.2.4)$$

其中 n_i $(i=1, 2, \cdots, s)$ 为正整数，$\sum_{i=1}^{s} n_i = n$，λ_i，$i=1, \cdots, s$ 为 A 的特征值，$J_{n_i}(\lambda_i)$ 为 $n_i \times n_i$ 阵

$$J_{n_i}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix},$$

称为 A 的 Jordan 块．若不计（2.2.4）中 Jordan 块的排列次序，则 A 的 Jordan 标准形是唯一的．

在一定意义下，可以认为对角阵是最简单的矩阵．但定理 2.2.2 表明，并不是每个方阵都可以相似于对角阵．但是如果对方阵加上一些条件，则它可以相似于对角阵．下面我们讨论一类可相似于对角阵的矩阵——正规矩阵．

定义 2.2.3 对 n 阶复方阵 A 和 B，若存在 n 阶酉阵 U，使得 $B=U^{*}AU$，则称 B 酉相似于 A．

定义 2.2.4 若复方阵 A 满足 $A^{*}A=AA^{*}$，则称 A 为正规矩阵．

Hermite 阵、酉阵、实对称阵和正交阵都是正规阵．

下面的定理刻画了正规阵的最重要的性质，其证明可在一般线性代数教程中找到．

定理 2.2.3 复方阵 A 酉相似于对角阵，当且仅当 A 是正规矩阵．

推论 2.2.1 (a) 设 A 为 $n\times n$ 正规阵，则 A 可表为

$$A=U\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}U^*, \qquad (2.2.5)$$

其中 U 为 $n\times n$ 酉阵，其列为 A 的标准正交化特征向量，λ_1，…，λ_n 为 A 的特征值.

(b) 若 A 为 Hermite 阵，则 (2.2.5) 中的 λ_1，…，λ_n 皆为实数.

(c) 若 A 为实对称阵，则 (2.2.5) 中的 λ_1，…，λ_n 皆为实数，且 U 为正交阵.

Hermite 阵和实对称阵是两类最重要的正规阵. 下面讨论它们的几个重要性质.

设 A 为 n 阶 Hermite 阵，若对任意 $x\in C^n$，有 $x^*Ax\geqslant 0$，则称 A 为半正定阵，记为 $A\geqslant 0$. 进一步，若 $x^*Ax=0$ 当且仅当 $x=0$，则称 A 为正定阵，记为 $A>0$.

定理 2.2.4 设 A 为 n 阶 Hermite 阵，则 $A\geqslant 0$ 当且仅当下列条件之一成立：

(a) A 的所有特征值非负；

(b) 对任一矩阵 Q，$Q^*AQ\geqslant 0$；

(c) 若 $r(A)=r$，存在 $r\times n$ 矩阵 B，使得 $A=B^*B$；

(d) 存在一个 Hermite 阵 B，使得 $A=B^2$.

类似地，对正定阵，我们有下面的定理.

定理 2.2.5 设 A 为 n 阶 Hermite 阵，则 $A>0$ 当且仅当下列条件之一成立：

(a) A 的所有特征值为正数；

(b) 对任一可逆复方阵 Q，$Q^*AQ>0$；

(c) 存在可逆复方阵 B，使得 $A=B^*B$.

如果 A 为 n 阶实对称阵，且对任意 $x\in R^n$，有 $x'Ax\geqslant 0$，则称 A 为半正定阵，记为 $A\geqslant 0$. 若 $A\geqslant 0$，且 $x'Ax=0$ 当且仅当 $x=0$，则称 A 为正定阵，记为 $A>0$. 对实对称半正定阵和正定阵，将定理 2.2.4 和 2.2.5 作明显的修改，相应的结论仍成立.

在结束这一节的时候，我们简要讨论一下相合关系.

定义 2.2.5 (a) 对两个 n 阶实方阵 A 和 B，若存在一个可逆实方阵 P，使得 $B=PAP'$，则称 B 与 A 为实相合的.

(b) 对两个 n 阶复方阵 A 和 B，若存在一个可逆复方阵 P，使得 $B=PAP^*$，则称 B 与 A 为复相合的.

容易验证，实相合与复相合都是等价关系.

定理 2.2.6（相合标准形） (a) 对任一 n 阶实对称方阵 A，必存在实可逆阵 P，使得

$$P'AP = \text{diag}(I_p, -I_{r-p}, 0),$$

其中 $r=r(A)$，即 A 的秩.

(b) 对任一 n 阶 Hermite 阵 A，必存在复可逆阵 P，使得

$$P^*AP = \text{diag}(I_p, -I_{r-p}, 0),$$

其中 $r=r(A)$.

注 在定理 2.2.6 中，p 为 A 的正特征值个数，$r-p$ 为 A 的负特征值个数，称 $2p-r$ 为 A 的符号差. 上面的定理表明，秩和符号差是相合关系下的不变量.

推论 2.2.2 设 A 为 $n\times n$ 半正定 Hermite 阵，其秩为 r，则存在可逆复方阵 P，使得

$$A = P^*\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}P.$$

§2.3 矩阵同时对角化

本节给出用一个可逆阵将两个 Hermite 阵或实对称阵同时对角化的一些很有用的充要条件.

定理 2.3.1 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为两个 n 阶 Hermite 阵，且 $\boldsymbol{B}>0$. 则存在可逆阵 $\boldsymbol{Q}$，使得

$$\boldsymbol{A}=\boldsymbol{Q}^*\boldsymbol{\Lambda}\boldsymbol{Q}, \tag{2.3.1}$$

$$\boldsymbol{B}=\boldsymbol{Q}^*\boldsymbol{Q}, \tag{2.3.2}$$

其中 $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，$\lambda_i$，$i=1, \cdots, n$ 为 $\boldsymbol{AB}^{-1}$的特征值.

证明 因 $\boldsymbol{B}>0$，故 $(\boldsymbol{B}^{\frac{1}{2}})^{-1}$存在（关于半正定阵的开方，详见定理 2.4.1）. 记

$$\boldsymbol{B}^{-1/2}=(\boldsymbol{B}^{\frac{1}{2}})^{-1},$$

则 $\boldsymbol{B}^{-1/2}>0$. 依定理 2.2.5 (b) 知

$$\boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2} \text{ 为 Hermite 阵.}$$

再应用推论 2.2.1 知，存在酉阵 $\boldsymbol{U}$，使得

$$\boldsymbol{U}^*(\boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2})\boldsymbol{U}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n), \tag{2.3.3}$$

这里 $\lambda_1, \cdots, \lambda_n$ 为 $\boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2}$的特征值. 也就是 $\boldsymbol{AB}^{-1}$的特征值，记 $\boldsymbol{Q}=\boldsymbol{U}^*\boldsymbol{B}^{1/2}$，从 (2.3.3) 可得 (2.3.1)，并易验证 (2.3.2) 成立. 定理证毕.

注 并不是任意两个 Hermite 阵都可以像定理 2.3.1 所述的那样用一个可逆阵将它们对角化. 例如

$$\boldsymbol{A}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \quad \boldsymbol{B}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

设

$$P=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

为任一可逆阵，则

$$P^* AP=\begin{pmatrix} a\bar{a} & \bar{a}b \\ a\bar{b} & b\bar{b} \end{pmatrix},$$

$$P^* BP=\begin{pmatrix} a\bar{a}+c\bar{c} & \bar{a}d+b\bar{c} \\ a\bar{d}+\bar{b}c & b\bar{d}+\bar{b}d \end{pmatrix}.$$

我们将证明 $\bar{a}b=0$ 与 $\bar{a}b+b\bar{c}=0$ 不能同时成立. 事实上，若 $\bar{a}b=0$，则从 P 为可逆阵知，或者 $a=0$，$b\neq0$，或者 $a\neq0$，$b=0$. 在第一种情况下，由

$$\bar{a}d+b\bar{c}=b\bar{c}=0,$$

必有 $c=0$. 同理，在第二种情形下可推出 $d=0$. 于是我们要使 P^*AP 与 P^*BP 同时为对角阵，必有 $a=c=0$ 或 $b=d=0$. 即在任何一种情况下，P 都是奇异阵.

定理 2.3.2 设 A、B 为同阶 Hermite 阵，且 A 可逆，则存在可逆阵 Q，使得 Q^*AQ 与 Q^*BQ 同时为对角阵，当且仅当 $A^{-1}B$ 相似于对角阵，且它的特征值都是实数.

证明 必要性 设 Q 为可逆阵，使得

$$Q^*AQ=D_1,$$

$$Q^*BQ=D_2,$$

其中 D_1 和 D_2 皆为对角阵. 于是

$$A^{-1}B=QD_1^{-1}D_2Q^{-1},$$

这就表明 $A^{-1}B$ 相似于对角阵 $D_1^{-1}D_2$. 注意到 D_1 和 D_2 的对角元都是实数，所以 $D_1^{-1}D_2$ 的对角元也皆为实数，且它们都是 $A^{-1}B$ 的特征值. 必要性得证.

充分性 设 $\lambda_1,\cdots,\lambda_k$ 为 $A^{-1}B$ 的 k 个不同特征值，其重

数分别为n_i，$i=1$，…，k，$n_1+\cdots+n_k=n$，且它们都是实数，T为可逆阵，使得

$$A^{-1}B = TDT^{-1}, \tag{2.3.4}$$

其中

$$D = \mathrm{diag}(\lambda_1 I_{n_1},\cdots,\lambda_k I_{n_k}).$$

从（2.3.4）有$BT=ATD$，从而

$$T^*BT = (T^*AT)D. \tag{2.3.5}$$

将T^*BT和T^*AT剖分成跟D相对应的块，即

$$T^*BT=\begin{pmatrix} G_{11} & \cdots & G_{1k} \\ \vdots & & \vdots \\ G_{k1} & \cdots & G_{kk} \end{pmatrix},$$

$$T^*AT=\begin{pmatrix} F_{11} & \cdots & F_{1k} \\ \vdots & & \vdots \\ F_{k1} & \cdots & F_{kk} \end{pmatrix}, \tag{2.3.6}$$

其中G_{ii}和F_{ii}为$n_i\times n_i$阵，$i=1,\cdots,k$. 因为A和B为Hermite阵，故有

$$G_{ij} = G_{ji}^*, \quad F_{ij} = F_{ji}^*, \quad i,j = 1,2,\cdots,k.$$

另外，从（2.3.5）可推出

$$G_{ij} = \lambda_j F_{ij},$$

于是

$$\lambda_j F_{ij} = G_{ij} = G_{ji}^* = \bar{\lambda}_i F_{ji}^* = \lambda_i F_{ij}.$$

因为λ_1，…，λ_k互不相同，于是上式表明

$$F_{ij} = 0, \quad 当\ i \neq j.$$

这时（2.3.6）变为

$$T^*AT = \mathrm{diag}(F_{11},\cdots,F_{kk}),$$

同时

$$T^*BT = \mathrm{diag}(\lambda_1 F_{11}, \cdots, \lambda_k F_{kk}).$$

注意到每个 F_{ii} 皆为 Hermite 阵，故存在酉阵 U_i，$i=1, \cdots, k$，使得

$$F_{ii} = U_i D_i U_i^*, i = 1, \cdots, k,$$

其中 D_i 皆为对角阵. 定义

$$U = \mathrm{diag}(U_1, \cdots, U_k),$$
$$Q = TU,$$

则有

$$Q^*AQ = \mathrm{diag}(D_1, \cdots, D_k),$$
$$Q^*BQ = \mathrm{diag}(\lambda_1 D_1, \cdots, \lambda_k D_k).$$

充分性得证. 定理证毕.

定理 2.3.3 设 A 和 B 为 n 阶 Hermite 阵，且不存在非零向量 x，使得

$$x^*Ax = x^*Bx = 0.$$

则存在可逆阵 Q，使得 Q^*AQ 与 Q^*BQ 皆为对角阵.

注意到，如果一个 Hermite 阵是上（下）三角阵，它一定是对角阵，于是定理 2.3.3 是下面定理的特殊情形：

定理 2.3.4 设 A 和 B 为两个 n 阶复方阵，且不存在非零向量 x，使得

$$x^*Ax = x^*Bx = 0. \tag{2.3.7}$$

则存在可逆阵 Q，使得 Q^*AQ 与 Q^*BQ 为上三角矩阵.

证明 对阶数 n 施以数学归纳法. $n=1$，命题显然成立.

现假设对阶数 $n-1$，命题成立. 若 A、B 之中有一个为可逆阵. 譬如说 B 可逆，则 $\det(A-\lambda B)$ 为一非常数多项式. 若 A 与 B 皆不可逆，则显然 $\det(A-\lambda B)=0$ 对 $\lambda=0$ 成立. 于是在任何一种情况下，$\det(A-\lambda B)=0$ 都有根 λ，因而存在非零向量 x_1，使得

$$Ax_1 = \lambda Bx_1. \tag{2.3.8}$$

根据假设(2.3.7)知，$Bx_1 \neq 0$. 故可选择 $n-1$ 个向量 $x_2, \cdots, x_n$，使得

$$x_i^* Bx_1 = 0, \quad i = 2, \cdots, n.$$

再利用(2.3.8)，有

$$x_i^* Ax_1 = 0, \quad i = 2, \cdots, n.$$

定义 $D = (x_1, \cdots, x_n)$，则

$$D^* AD = \left(\begin{array}{c|c} x_1^* Ax_1 & \cdots \quad x_1^* Ax_n \\ \hline 0 & \\ \vdots & A_1 \\ 0 & \end{array}\right),$$

$$D^* BD = \left(\begin{array}{c|c} x_1^* Bx_1 & \cdots \quad x_1^* Bx_n \\ \hline 0 & \\ \vdots & B_1 \\ 0 & \end{array}\right).$$

现在我们证明 D 是可逆阵. 用反证法. 设

$$x_1 = c_2 x_2 + \cdots + c_n x_n,$$

则

$$x_1^* Ax_1 = (\bar{c}_2 x_2^* + \cdots + \bar{c}_n x_n^*) Ax_1 = 0,$$

类似地，可证

$$x_1^* Bx_1 = 0,$$

这与假设(2.3.7)矛盾. 故 D 是可逆阵.

下面我们证明：子矩阵 A_1 和 B_1 仍满足定理假设，即不存在非零向量 $u \in C^{n-1}$，使得

$$u^* A_1 u = u^* B_1 u = 0.$$

用反证法. 设 $u=(u_2, \cdots, u_n)'\neq 0$，满足

$$u^* A_1 u = u^* B_1 u = 0. \tag{2.3.9}$$

记 $D_1=(x_2, \cdots, x_n)$，容易验证

$$A_1 = D_1^* A D_1, \tag{2.3.10}$$

$$B_1 = D_1^* B D_1, \tag{2.3.11}$$

于是，对 $y=D_1u=u_2x_2+\cdots+u_nx_n\neq 0$（注意，$y\neq 0$ 来自 D 的可逆性），从（2.3.9）—（2.3.11）可推得

$$y^* Ay = y^* By = 0,$$

这与假设（2.3.7）相矛盾.

因为 A_1 和 B_1 仍满足定理假设，再根据归纳假设，存在 $n-1$ 阶可逆阵 Q_1 使得 $Q_1^* A_1 Q_1$ 与 $Q_1^* BQ_1$ 皆为上三角阵. 最后令

$$Q = D\begin{pmatrix} 1 & 0 \\ 0 & Q_1 \end{pmatrix},$$

则 $Q^* AQ$ 与 $Q^* BQ$ 皆为上三角阵. 定理证毕.

定理 2.3.5 设 A 与 B 皆为 n 阶正规阵. 则存在酉阵 U，使得 $U^* AU$ 与 $U^* BU$ 同时为对角阵的充要条件是 $AB=BA$.

证明 必要性的证明是容易的. 下面证明充分性.

因为 A 为正规阵，依定理 2.2.3，存在酉阵 T，使得 $T^* AT$ 为对角阵. 若 $\lambda_1, \cdots, \lambda_k$ 为 A 的 k 个互不相同的特征值，其重数分别为 $n_1, \cdots, n_k$，则

$$T^* AT = \mathrm{diag}(\lambda_1 I_{n_1}, \cdots, \lambda_k I_{n_k}). \tag{2.3.12}$$

由 $AB=BA$ 得

$$T^* AT \cdot T^* BT = T^* BTT^* AT. \tag{2.3.13}$$

记 $A_1=T^* AT$，$B_1=T^* BT$，则上式即为

$$A_1B_1 = B_1A_1.$$

因为 A_1 具有 (2.3.12) 的形式，故从上式知 B_1 必为如下准对角阵形式：

$$B_1 = \mathrm{diag}(C_1, \cdots, C_k),$$

其中 C_i 为 n_i 阶方阵，$i=1, \cdots, k$. 因为 $B^*B=BB^*$ 可推出 $B_1^*B_1=B_1B_1^*$，故

$$C_i^*C_i = C_iC_i^*, \qquad i = 1, \cdots, k,$$

于是 C_i，$i=1, \cdots, k$ 都是正规阵. 因而对每个 C_i，都存在 n_i 阶酉阵 Q_i，使得

$$Q_i^*C_iQ_i = \mathrm{diag}(\mu_{i1}, \cdots, \mu_{in_i}), \quad i = 1, \cdots, k.$$

记 $Q=\mathrm{diag}(Q_1, \cdots, Q_k)$，则 Q 为酉阵. 再令 $U=TQ$，它仍为酉阵，且

$$U^*AU = \mathrm{diag}(\lambda_1 I_{n_1}, \cdots, \lambda_k I_{n_k}),$$

$$\begin{aligned} U^*BU &= Q^*\mathrm{diag}(C_1, \cdots, C_k) \\ &= \mathrm{diag}(\mu_{11}, \cdots, \mu_{1n_1}, \cdots, \mu_{k1}, \cdots, \mu_{kn_k}). \end{aligned}$$

定理证毕.

注 定理 2.3.5 表明，若两个 n 阶正规阵 A 和 B 是可交换的，则它们有 n 个相互正交的公共特征向量.

推论 2.3.1 设 A 与 B 为同阶 Hermite 阵，则存在酉阵 U，使得 U^*AU 和 U^*BU 为对角阵，当且仅当 $AB=BA$.

推论 2.3.2 设 A、B 为同阶实对称阵，则存在正交阵 P，使得 $P'AP$ 与 $P'BP$ 为对角阵，当且仅当 $AB=BA$.

定理 2.3.6 设 A、B 为 n 阶半正定 Hermite 阵，则存在可逆复方阵 Q，使得 Q^*AQ 与 Q^*BQ 皆为对角阵.

证明 设 $t=r(A+B)$. 由推论 2.2.2，存在可逆复方阵 P，使得

$$P^*(A+B)P = \begin{pmatrix} I_t & 0 \\ 0 & 0 \end{pmatrix}. \qquad (2.3.14)$$

将 P^*BP 分块为

$$P^*BP = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \qquad (2.3.15)$$

其中 C_{11} 为 $t\times t$ 矩阵. 用 $M\geqslant N$ 表示 $M-N\geqslant 0$，则因

$$P^*(A+B)P \geqslant P^*BP,$$

结合（2.3.14）和（2.3.15），得

$$C_{ij} = 0, \qquad i+j\geqslant 3,$$

即

$$P^*BP = \begin{pmatrix} C_{11} & 0 \\ 0 & 0 \end{pmatrix}. \qquad (2.3.16)$$

因为 C_{11} 仍为 Hermite 阵，从推论 2.2.1（b）知，存在 t 阶酉阵 Q_1，使得

$$Q_1^*C_{11}Q_1 = \mathrm{diag}(\lambda_1,\cdots,\lambda_t).$$

记

$$Q = P\begin{pmatrix} Q_1 & 0 \\ 0 & I_{n-t} \end{pmatrix},$$

这是一个可逆方阵，且使得

$$Q^*BQ = \mathrm{diag}(\lambda_1,\cdots,\lambda_t,0,\cdots,0),$$

$$Q^*AQ = \mathrm{diag}(1-\lambda_1,\cdots,1-\lambda_t,0,\cdots,0).$$

证毕.

推论 2.3.3 设 A 和 B 为同阶半正定实对称阵. 则存在实可逆阵 Q，使得 $Q'AQ$ 与 $Q'BQ$ 皆为对角阵.

§2.4 矩 阵 分 解

所谓矩阵分解，就是将一个矩阵分解为从某种意义上讲

比较简单或对它的性质比较熟悉的若干个矩阵的乘积．在前两节，我们讨论了矩阵在相抵、相似和相合关系下的标准形以及两个矩阵的同时对角化，当然这些内容也可以概括在"矩阵分解"这一标题之下．之所以把它们分开讨论，一方面是由于两者之间还存在着一定的差异，另一方面是为了技术处理上的方便．

下面是矩阵分解的一些最常用的结果．

定理 2.4.1（半正定阵的开方）　设 $\boldsymbol{A}$ 为半正定（正定）Hermite 阵．则存在唯一半正定（正定）Hermite 阵 $\boldsymbol{B}$，使得

$$\boldsymbol{A}=\boldsymbol{B}^2. \tag{2.4.1}$$

证明　只证明 $\boldsymbol{A}\geqslant 0$ 的情形．设 $\boldsymbol{A}$ 为 n 阶方阵，则存在酉阵 $\boldsymbol{U}$，使得

$$\boldsymbol{A}=\boldsymbol{U}\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\boldsymbol{U}^*,$$

其中 $\lambda_1\geqslant\cdots\geqslant\lambda_n\geqslant 0$ 为 $\boldsymbol{A}$ 的特征值．令

$$\boldsymbol{B}=\boldsymbol{U}\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{U}^*, \tag{2.4.2}$$

显然 $\boldsymbol{B}\geqslant 0$ 且 $\boldsymbol{B}^2=\boldsymbol{A}$．下证唯一性．假设存在另一个 Hermite 阵 $\boldsymbol{C}\geqslant 0$，满足 $\boldsymbol{A}=\boldsymbol{C}^2$，那么存在一个酉阵 $\boldsymbol{W}$，使得

$$\boldsymbol{C}=\boldsymbol{W}\mathrm{diag}(\mu_1,\cdots,\mu_n)\boldsymbol{W}^*,$$

这里 $\mu_1\geqslant\cdots\geqslant\mu_n\geqslant 0$．因

$$\boldsymbol{C}^2=\boldsymbol{W}\mathrm{diag}(\mu_1^2,\cdots,\mu_n^2)\boldsymbol{W}^*=\boldsymbol{A},$$

故 $\mu_i^2=\lambda_i$，$i=1,\cdots,n$，即 $\mu_i=\lambda_i^{1/2}$．于是

$$\boldsymbol{C}=\boldsymbol{W}\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{W}^*.$$

再从 $\boldsymbol{A}=\boldsymbol{B}^2=\boldsymbol{C}^2$ 可得

$$\boldsymbol{U}\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\boldsymbol{U}^*=\boldsymbol{W}\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\boldsymbol{W}^*,$$

等价地，

$$\boldsymbol{W}^*\boldsymbol{U}\mathrm{diag}(\lambda_1,\cdots,\lambda_n)=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\boldsymbol{W}^*\boldsymbol{U}.$$

记 $W^*U=(d_{ij})$. 则有

$$d_{ij}\lambda_j=\lambda_i d_{ij},\quad i,j=1,\cdots,n.$$

因而当 $\lambda_i\neq\lambda_j$ 时，必有 $d_{ij}=0$. 因此无论 λ_i 与 λ_j 是否相等，我们总有

$$d_{ij}\lambda_j^{1/2}=d_{ij}\lambda_i^{1/2},\quad i,j=1,\cdots,n,$$

也就是

$$W^*U\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})=\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})W^*U,$$

等价地，

$$U\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})U^*=W\mathrm{diag}(\lambda_1^{1/2},\cdots,\lambda_n^{1/2})W^*.$$

唯一性得证，定理证毕.

注 我们称(2.4.2)定义的 B 为 A 的平方根，常常记为 $A^{1/2}$.

定理 2.4.2 (奇异值分解) 设 A 为 $m\times n$ 秩为 r 的复(实)矩阵. 则存在 m 阶酉(正交)阵 P 和 n 阶酉(正交)阵 Q，使得

$$A=P\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix}Q^*,\qquad (2.4.3)$$

其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_r),\lambda_i>0$，$i=1,\ \cdots,\ r$.

证明 显然 $A^*A\geqslant0$，于是它的特征值皆非负，记 $\lambda_1^2,\cdots\lambda_r^2$ 为 A^*A 的正特征值. 依推论 2.2.1，存在酉阵 Q，使得

$$Q^*A^*AQ=\begin{pmatrix}\Lambda^2 & 0\\ 0 & 0\end{pmatrix},$$

这里 $\Lambda^2=\mathrm{diag}(\lambda_1^2,\cdots,\lambda_r^2)$. 记 $B=AQ$，则上式变为

$$B^*B=\begin{pmatrix}\Lambda^2 & 0\\ 0 & 0\end{pmatrix},$$

这表明 B 的 n 个列互相正交，且前 r 个列向量长度分别为 λ_1，$\cdots$，λ_r (这里 $\lambda_i>0$ 分别为 λ_i^2 的平方根). 于是存在一酉阵 P，

使得

$$B = P\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix},$$

这里 $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_r)$. 从上式和 B 的定义，可得 (2.4.3)，定理证毕.

注 1 从定理证明过程知，在 A 的奇异值分解 (2.4.3) 中，Q 的列向量为 A^*A 的标准正交化特征向量，λ_1^2，…，λ_r^2 为 A^*A 或 AA^* 的正特征值. 对应地，P 的列向量为 AA^* 的标准正交化特征向量.

注 2 λ_1，…，λ_r 称为 A 的奇异值.

奇异值分解是矩阵最有用的分解形式之一，它在数值分析、近似计算、解线性方程组和广义逆研究中具有广泛的应用，特别，它是研究 Moore-Penrose 逆的最重要的工具之一.

众所周知，任何一个复数 z，均可表为 $z=|z|e^{i\varphi}$，这里右端第一因子 $|z|\geqslant 0$，而第二因子的模为 1. 类似于这个初等事实，我们有下面的矩阵极分解.

定理 2.4.3（极分解） 设 A 为 n 阶复（实）方阵. 则存在 n 阶酉（正交）阵 U 和 n 阶半正定 Hermite（实对称）阵 S，使得

$$A = SU. \tag{2.4.4}$$

若 A 可逆，则这种分解是唯一的.

证明 从奇异值分解知，存在酉阵 P 和 Q，使得

$$A = P\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q^*,$$

这里 $\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_r)$，$\lambda_i>0$，$i=1,\cdots,r$，$r=r(A)$. 将上式改写为

$$A = P\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} P^* PQ^*$$

$$\stackrel{d}{=} SU,$$

其中

$$S = P\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} P^* \geqslant 0,$$

$$U = PQ^*,$$

且 U 为酉阵. (2.4.4) 得证.

现在证明当 A 可逆时分解是唯一的. 假设 $A = S_1U_1 = S_2U_2$，其中 U_1 和 U_2 皆为酉阵，Hermite 阵 S_1 和 S_2 皆正定. 从

$$AA^* = S_1^2 = S_2^2$$

知 S_1 和 S_2 均为 AA^* 的平方根阵. 依定理 2.4.1 有 $S_1 = S_2$. 另一方面

$$U_1 = S_1^{-1}A = S_2^{-1}A = U_2,$$

唯一性得证. 定理证毕.

定理 2.4.4 (Schur 酉三角分解) 设 A 为复方阵，则存在酉阵 U，使得

$$A = ULU^*, \tag{2.4.5}$$

其中 L 为上（或下）三角阵，其对角元为 A 的特征值. 若 A 为实方阵且特征值也是实的，则 U 为正交阵.

证明 对 A 的阶数施行数学归纳法. 对 $n=1$，命题显然成立.

现假定命题对 $n-1$ 阶方阵成立. 设 A 为 n 阶方阵，其特征值为 λ_1，…，λ_n，x_1 为 A_1 的对应于 λ_1 的单位特征向量. 将 x_1 扩充为 C^n 的一组基：x_1，y_2，…，y_n. 应用 Schmidt 正交化程序，将这些向量变换为 C^n 的一组标准正交基：x_1，z_2，…，

z_n. 定义酉阵 $U_0=(x_1,z_2,\cdots,z_n)$，则

$$U_0^* A U_0 = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix},$$

其中 A_1 为 $n-1$ 阶方阵，且特征值为 λ_2，…，λ_n. 依归纳法假设，存在 $n-1$ 阶酉阵 U_1，使得

$$A_1 = U_1 L_1 U_1^*,$$

这里 L_1 为上三角阵，其对角元为 A_1 的特征值 λ_2，…，λ_n. 令

$$U = U_0 \begin{pmatrix} 1 & 0 \\ 0 & U_1 \end{pmatrix},$$

$$L = \begin{pmatrix} \lambda_1 & * \\ 0 & L_1 \end{pmatrix},$$

其中 U 为酉阵，易验证对这样的 U 和 L（2.4.5）成立. 这就证明了 L 为上三角阵的情形.

若对 A^* 作上面的分解：$A^*=ULU^*$，这里 U 为酉阵，L 为上三角阵，则 $A=UL^*U^*$，此时 L 就是下三角阵了.

最后，当 A 为实方阵，且所有特征值皆为实数，则它的特征向量都可取为实向量，此时酉阵便为正交阵. 定理证毕.

定理 2.4.5（满秩分解） 设 A 为 $m\times n$ 复阵，$r=r(A)$. 则存在 $m\times r$ 和 $r\times n$ 且秩皆为 r 的矩阵 B 和 C，使得 $A=BC$.

证明 根据定理 2.2.1，存在 $m\times m$ 和 $n\times n$ 可逆阵 P 和 Q，使得

$$A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q. \tag{2.4.6}$$

对 P 和 Q 作分块：

$$P = (P_1 \vdots P_2), \quad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}, \tag{2.4.7}$$

其中 P_1 为 $m\times r$ 阵，Q_1 为 $r\times n$ 阵，且 $r(P)=r(Q)=r$. 综合(2.4.6)和(2.4.7)，便有 $A=P_1Q_1$，证毕.

定理 2.4.6 (*QR* 分解) 设 A 为 $m\times n$ 阵，$r(A)=n$. 则存在 $m\times n$ 阵 Q 和 $n\times n$ 上三角阵 R，使得 $A=QR$，其中 $Q^*Q=I_n$，R 的对角元为正数. 若 A 为实矩阵，则 Q 和 R 也是实矩阵.

证明 对矩阵 A 的列施以 Schmidt 正交化程序，便可得到所要结论.

注 若 $r(A)=r<n$，则分解式 $A=QR$ 仍然成立，且 $Q^*Q=I_n$，但此时 R 的一些对角元可以为零（参阅 Horn 和 Johnson 1985，p. 112）.

定理 2.4.7 (Cholesky 分解) 设 A 为 $n\times n$ 正定 Hermite 阵. 则存在上三角阵 L，其对角元皆为正数，使得 $A=L^*L$. 若 A 为正定实对称阵，则 L 的元素皆为实数.

证明 因 $A>0$，故存在 $B>0$，使得 $A=B^*B$. 对 B 作 *QR* 分解：$B=QR$，这里 Q 为酉阵，R 为上三角阵，其对角元皆为正数. 于是

$$A=B^*B=R^*Q^*QR=R^*R.$$

利用定理 2.4.6，剩余部分的证明是显然的. 定理证毕.

对于可对角化方阵，还有一种分解，称为谱分解，将留在 §2.7 讨论.

§2.5 Schur 补

设 A 为 n 阶方阵，将其分块为

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}, \tag{2.5.1}$$

这里 A_{11} 为 t 阶方阵. 若 A_{11} 可逆，则称 $(n-t)$ 阶方阵

$$A_{22}-A_{21}A_{11}^{-1}A_{12}$$

为 A_{11} 在 A 中的 Schur 补(Schur complement). 若 A_{11} 不可逆，则称

$$A_{22}-A_{21}A_{11}^{-}A_{12}$$

为 A_{11} 在 A 中的广义 Schur 补. 这两种 Schur 补都简记为 $A_{22\cdot 1}$ 或 (A/A_{11}). 记号 $A_{22\cdot 1}$ 在数理统计学中被广泛使用，但 (A/A_{11}) 也有其方便之处. 在后面的讨论中，我们基本上都使用 $A_{22\cdot 1}$，有时迫不得已也使用 (A/A_{11}). 不难明白，Schur 补在分块矩阵的广义逆研究中占有重要的地位.

用 $\det A$ 表示方阵 A 的行列式，记号

$$\operatorname{In}A=(\pi,\nu,\delta)$$

表示 A 的惯性指数(inertia)，这里 π，ν 和 δ 分别表示 A 的正、负和零特征值的个数，它们都是相合关系下的不变量.

定理 2.5.1 设 n 阶方阵 A 有 (2.5.1) 的形式，且 A_{11} 可逆. 则

(a) $\qquad \det A=\det A_{11}\det A_{22\cdot 1}$;

(b) $\qquad r(A)=r(A_{11})+r(A_{22\cdot 1})$;

(c) $\qquad \operatorname{In}A=\operatorname{In}A_{11}+\operatorname{In}A_{22\cdot 1}$.

证明 因为

$$B=\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

$$=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22\cdot 1} \end{pmatrix}, \qquad (2.5.2)$$

于是前两条是显然的. 利用定理 2.2.6 容易证明

$$\operatorname{In}A=\operatorname{In}CAC^{*},$$

这里 C 为可逆阵. 于是

$$\mathrm{In}A = \mathrm{In}B = \mathrm{In}A_{11} + \mathrm{In}A_{22\cdot1}.$$

定理证毕.

类似地，我们有

推论 2.5.1 设 A 有 (2.5.1) 形式，A_{22} 可逆，则

(a) $\det A = \det A_{22} \det A_{11\cdot2}$;

(b) $r(A) = r(A_{22}) + r(A_{11\cdot2})$;

(c) $\mathrm{In}A = \mathrm{In}A_{22} + \mathrm{In}A_{11\cdot2}$.

从 (2.5.2)，我们很容易证明如下定理.

定理 2.5.2 设 A 为 Hermite 阵.

(a) 则 $A>0 \Leftrightarrow A_{11}>0$，$A_{22\cdot1}>0$;

(b) 若 $A_{11}>0$，则 $A\geqslant0 \Leftrightarrow A_{22\cdot1}\geqslant0$.

定理 2.5.3 设 A 和 B 为同阶半正定 Hermite 阵且分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

又 A_{11} 和 B_{11} 为可逆阵，则

$$(A+B)_{22\cdot1} \geqslant A_{22\cdot1} + B_{22\cdot1}.$$

证明 在矩阵恒等式

$$(E - FH^{-1}G)^{-1} = E^{-1} + E^{-1}F(H - GE^{-1}F)^{-1}GE^{-1}$$

中，令 $E=F=G=A_{11}$，$H=-B_{11}$，得

$$(A_{11} + A_{11}B_{11}^{-1}A_{11})^{-1} = A_{11}^{-1} - (A_{11} + B_{11})^{-1}. \tag{2.5.3}$$

另一方面

$$\begin{aligned} \Delta &= (A+B)_{22\cdot1} - A_{22\cdot1} - B_{22\cdot1} \\ &= A_{21}A_{11}^{-1}A_{12} + B_{21}B_{11}^{-1}B_{12} \end{aligned}$$

$$-(A_{21}+B_{21})(A_{11}+B_{11})^{-1}(A_{12}+B_{12}).$$

利用（2.5.3），我们有

$$\begin{aligned}\Delta &= (A_{21} \vdots B_{21})\begin{pmatrix} A_{11}^{-1}-(A_{11}+B_{11})^{-1} & -(A_{11}+B_{11})^{-1} \\ -(A_{11}+B_{11})^{-1} & B_{11}^{-1}-(A_{11}+B_{11})^{-1}\end{pmatrix}\begin{pmatrix} A_{12} \\ B_{12}\end{pmatrix} \\ &= (A_{21} \vdots B_{21})\begin{pmatrix} I \\ -B_{11}^{-1}A_{11}\end{pmatrix}(A_{11}+A_{11}B_{11}^{-1}A_{11})^{-1}(I \vdots -A_{11}B_{11}^{-1})\begin{pmatrix} A_{12} \\ B_{12}\end{pmatrix} \\ &\geqslant 0.\end{aligned}$$

定理证毕.

考虑分块矩阵

$$A=\begin{pmatrix} B & C \\ D & E\end{pmatrix}=\left(\begin{array}{cc:c} B_{11} & B_{12} & C_1 \\ B_{21} & B_{22} & C_2 \\ \hdashline D_1 & D_2 & E\end{array}\right), \qquad (2.5.4)$$

其中

$$B=\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{pmatrix}, \quad C=\begin{pmatrix} C_1 \\ C_2\end{pmatrix}, \quad D=(D_1, D_2).$$

这里我们有三个 Schur 补：

$$(A/B)、(A/B_{11})、(B/B_{11}).$$

下面的定理建立了这三者之间的一个等式.

定理 2.5.4 在分块矩阵(2.5.4)中，设 B 和 B_{11} 皆可逆. 则

(a) Schur 补 (B/B_{11}) 是 Schur 补 (A/B_{11}) 的可逆主子阵；

(b) $$(A/B)=((A/B_{11})/(B/B_{11})). \qquad (2.5.5)$$

证明 因为

$$(A/B_{11})=\begin{pmatrix} B_{22} & C_2 \\ D_2 & E\end{pmatrix}-\begin{pmatrix} B_{21} \\ D_1\end{pmatrix}B_{11}^{-1}(B_{12} \vdots C_1)$$

$$= \begin{pmatrix} B_{22} - B_{21}B_{11}^{-1}B_{12} & C_2 - B_{21}B_{11}^{-1}C_1 \\ D_2 - D_1B_{11}^{-1}B_{12} & E - D_1B_{11}^{-1}C_1 \end{pmatrix}. \tag{2.5.6}$$

易见 $B_{22}-B_{21}B_{11}^{-1}B_{12}=(B/B_{11})$ 是 (A/B_{11}) 的主子阵. 由假设 B 和 B_{11} 皆可逆，由定理 2.5.1 (a) 知 (B/B_{11}) 可逆. 这就证明了 (a)。

将 (2.5.6) 改写为

$$(A/B_{11}) = \begin{pmatrix} (B/B_{11}) & C_2 - B_{21}B_{11}^{-1}C_1 \\ D_2 - D_1B_{11}^{-1}B_{12} & E - D_1B_{11}^{-1}C_1 \end{pmatrix}.$$

依 Schur 补的定义有

$$\begin{aligned}
&((A/B_{11})/(B/B_{11})) \\
&= E - D_1B_{11}^{-1}C_1 - (D_2 - D_1B_{11}^{-1}B_{12})(B/B_{11})^{-1}(C_2 - B_{21}B_{11}^{-1}C_1) \\
&= E - (D_1 \vdots D_2) \cdot \\
&\quad \begin{pmatrix} B_{11}^{-1} + B_{11}^{-1}B_{12}(B/B_{11})^{-1}B_{21}B_{11}^{-1} & -B_{11}^{-1}B_{12}(B/B_{11})^{-1} \\ -(B/B_{11})^{-1}B_{21}B_{11}^{-1} & (B/B_{11})^{-1} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \\
&= E - DB^{-1}C \\
&= (A/B),
\end{aligned}$$

这里我们应用了分块矩阵求逆公式：

$$\begin{aligned}
B^{-1} &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} B_{11}^{-1} + B_{11}^{-1}B_{12}(B/B_{11})^{-1}B_{21}B_{11}^{-1} & -B_{11}^{-1}B_{12}(B/B_{11})^{-1} \\ -(B/B_{11})^{-1}B_{21}B_{11}^{-1} & (B/B_{11})^{-1} \end{pmatrix}.
\end{aligned}$$

定理证毕.

注 等式 (2.5.5) 叫做 Schur 补的商性质 (quotient

property)（见 Ouellette (1981)）. 与此相平行的一个等式是

$$AB^{-1} = (AB_{11}^{-1})(BB_{11}^{-1})^{-1}.$$

本节讨论的都是 Schur 补. 关于广义 Schur 补将在 § 6.3 讨论.

§ 2.6 幂等阵与投影阵

定义 2.6.1 若 n 阶方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{A}$，则称 $\boldsymbol{A}$ 为幂等阵 (idempotent matrix).

幂等阵有下面的性质：

定理 2.6.1 设 $\boldsymbol{A}$ 为 n 阶幂等阵，则

(a) $\boldsymbol{A}^*$ 和 $\boldsymbol{I}-\boldsymbol{A}$ 都是幂等阵；

(b) $\boldsymbol{A}$ 的特征值只能是 0 或 1，且特征值 1 的重数为 $\boldsymbol{A}$ 的秩；

(c) $r(\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A})$，这里 $\mathrm{tr}(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 的迹，即 $\boldsymbol{A}$ 的对角线元素之和；

(d) $\mathscr{N}(\boldsymbol{A})=\mathscr{M}(\boldsymbol{I}-\boldsymbol{A})$；

(e) $\mathscr{M}(\boldsymbol{A})=\mathscr{N}(\boldsymbol{I}-\boldsymbol{A})$.

这些事实的证明都不难，留给读者作练习.

设 $\boldsymbol{C}^n$ 有直接和分解：

$$\boldsymbol{C}^n = S_1 \oplus S_2,$$

那末对任一向量 $\boldsymbol{x}\in\boldsymbol{C}^n$，可唯一地分解为

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}, \tag{2.6.1}$$

其中 $\boldsymbol{y}\in S_1$，$\boldsymbol{z}\in S_2$，我们称 $\boldsymbol{y}$ 为 $\boldsymbol{x}$ 循 S_2 在 S_1 上的投影. 从 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的变换是一个线性变换. 从一个有限维空间到另一个有限维空间的线性变换能够用一个矩阵来表示，我们称该矩阵为变换矩阵. 当选定这两个空间的基之后，线性变换与其变

换矩阵之间有一个一一对应关系，于是我们总是用同一个字母既表示线性变换又表示对应的变换矩阵. (2.6.1) 所定义的变换记之为 P_{S_1,S_2}，称为投影阵.

定理 2.6.2 P 为一个投影阵当且仅当 P 是幂等阵.

证明 设 P 为循 S_2 在 S_1 的投影阵，则由定义知，对任意的 $x\in S_1$，它循 S_2 在 S_1 的投影就是 x 本身. 于是

$$P^2x = PPx = Px,$$

对一切 $x\in C^n$ 成立，这就证明了 $P^2=P$.

反过来，设 P 为幂等阵. 定义

$$S_1 = \{y : y = Px, \forall\ x \in C^n\},$$

$$S_2 = \{z : Pz = 0\},$$

则 $S_1\cap S_2= \{0\}$. 对任给的 $x\in C^n$，可将其分解为

$$x = Px + (I - p)x,$$

这里 $Px\in S_1$，$(I-P)x\in S_2$，故

$$C^n = S_1 \oplus S_2.$$

这就证明了 P 为 P_{S_1,S_2}. 定理证毕.

注意到 $S_1=\mathscr{M}(P)$，$S_2=\mathscr{N}(P)$，因此，若 P 为幂等阵，必有

$$P = P_{\mathscr{M}(P),\mathscr{N}(P)}.$$

这就证明了如下推论：

推论 2.6.1 对任意幂等阵 A，都有

$$A = P_{\mathscr{M}(A),\mathscr{N}(A)}. \tag{2.6.2}$$

定理 2.6.3 设 S_1 和 S_2 为 C^n 的两个互补子空间，则存在唯一投影阵 P_{S_1,S_2}，使得

$$\mathscr{M}(P_{S_1,S_2}) = S_1, \tag{2.6.3}$$

$$\mathscr{N}(\boldsymbol{P}_{S_1,S_2}) = S_2. \tag{2.6.4}$$

证明 设 $\dim S_1 = g$，则 $\dim S_2 = n - g$. 记

$$\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_g\}$$

和

$$\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{n-g}\}$$

分别为 S_1 和 S_2 的基，则

$$\boldsymbol{P}_{S_1,S_2}\boldsymbol{a}_i = \boldsymbol{a}_i, \quad i = 1, \cdots, g, \tag{2.6.5}$$

$$\boldsymbol{P}_{S_1,S_2}\boldsymbol{b}_i = \boldsymbol{0}, \quad i = 1, \cdots, n-g. \tag{2.6.6}$$

记 $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_g)$, $\boldsymbol{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{n-g})$，则(2.6.5)和(2.6.6)等价于

$$\boldsymbol{P}_{S_1,S_2}(\boldsymbol{A} \vdots \boldsymbol{B}) = (\boldsymbol{A} \vdots \boldsymbol{0}). \tag{2.6.7}$$

因为矩阵 $(\boldsymbol{A} \vdots \boldsymbol{B})$ 为可逆阵，故上式有唯一解

$$\boldsymbol{P}_{S_1,S_2} = (\boldsymbol{A} \vdots \boldsymbol{0})(\boldsymbol{A} \vdots \boldsymbol{B})^{-1}. \tag{2.6.8}$$

这就证明了 $\boldsymbol{P}_{S_1,S_2}$ 的唯一性.

从 (2.6.8) 我们立即看出

$$\mathscr{M}(\boldsymbol{P}_{S_1,S_2}) = \mathscr{M}(\boldsymbol{A} \vdots \boldsymbol{0}) = \mathscr{M}(\boldsymbol{A}) = S_1.$$

另一方面，对任给的 $\boldsymbol{x} \in S_2$，必存在 $\boldsymbol{u}$，使得 $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{u}$. 于是

$$\boldsymbol{P}_{S_1,S_2}\boldsymbol{x} = \boldsymbol{P}_{S_1,S_2}\boldsymbol{B}\boldsymbol{u} = \boldsymbol{0},$$

这表明 $S_2 \subset \mathscr{N}(\boldsymbol{P}_{S_1,S_2})$. 但它们的维数皆为 $n-g$，这是因为 $\dim \mathscr{N}(\boldsymbol{P}_{S_1,S_2}) = n - r(\boldsymbol{P}_{S_1,S_2}) = n - g$，于是 $S_2 = \mathscr{N}(\boldsymbol{P}_{S_1,S_2})$. 定理证毕.

注 (2.6.8) 提供了投影阵的一种表示，但这种表示应用上有时很不方便. 后面我们将给出另外的表示式（见§3.6）.

两个投影阵的和、差、积未必是投影阵. 对此，我们有下面三个定理.

定理 2.6.4 (a) 设 $\boldsymbol{P}_1$ 和 $\boldsymbol{P}_2$ 为两个投影阵，则 $\boldsymbol{P}=\boldsymbol{P}_1+\boldsymbol{P}_2$ 为投影阵，当且仅当

$$\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{0}. \tag{2.6.9}$$

(b) 设 $\boldsymbol{P}_1=\boldsymbol{P}_{S_1,T_1}$，$\boldsymbol{P}_2=\boldsymbol{P}_{S_2,T_2}$. 若 (2.6.9) 成立，则

$$\boldsymbol{P}\overset{d}{=}\boldsymbol{P}_1+\boldsymbol{P}_2=\boldsymbol{P}_{S,T}, \tag{2.6.10}$$

其中 $S=S_1\oplus S_2, T=T_1\cap T_2$.

证明 (a) 充分性的证明是容易的，下证必要性. 因为 $\boldsymbol{P}_1^2=\boldsymbol{P}_1$，$\boldsymbol{P}_2^2=\boldsymbol{P}_2$，由 $\boldsymbol{P}^2=\boldsymbol{P}$ 可得

$$\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{0}. \tag{2.6.11}$$

用 $\boldsymbol{P}_1$ 左乘上式，我们有

$$\boldsymbol{P}_1(\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_2\boldsymbol{P}_1)=\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{0}, \tag{2.6.12}$$

再将上式右乘 $\boldsymbol{P}_1$，得

$$(\boldsymbol{P}_1\boldsymbol{P}_2+\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1)\boldsymbol{P}_1=2\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{0}.$$

结合 (2.6.12) 知，$\boldsymbol{P}_1\boldsymbol{P}_2=\boldsymbol{0}$. 再从 (2.6.11) 得 $\boldsymbol{P}_2\boldsymbol{P}_1=\boldsymbol{0}$. 这就完成了 (a) 的证明.

(b) 根据推论 2.6.1，知 (2.6.10) 等价于

$$\mathscr{M}(\boldsymbol{P}_1+\boldsymbol{P}_2)=S_1\oplus S_2, \tag{2.6.13}$$

$$\mathscr{N}(\boldsymbol{P}_1+\boldsymbol{P}_2)=T_1\cap T_2, \tag{2.6.14}$$

但是

$$\mathscr{M}(\boldsymbol{P}_1)=S_1,\quad \mathscr{N}(\boldsymbol{P}_1)=T_1,$$

$$\mathscr{M}(\boldsymbol{P}_2)=S_2,\quad \mathscr{N}(\boldsymbol{P}_2)=T_2,$$

于是结合 (2.6.13) 和 (2.6.14)，问题归结为证明

$$\mathscr{M}(\boldsymbol{P}_1+\boldsymbol{P}_2)=\mathscr{M}(\boldsymbol{P}_1)\oplus\mathscr{M}(\boldsymbol{P}_2), \tag{2.6.15}$$

$$\mathscr{N}(\boldsymbol{P}_1+\boldsymbol{P}_2)=\mathscr{N}(\boldsymbol{P}_1)\cap\mathscr{N}(\boldsymbol{P}_2). \tag{2.6.16}$$

对每个 $\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P}_1)$，有 $\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{x}$. 于是，利用 (2.6.9)，

得

$$Px = PP_1x = P_1^2x = P_1x = x,$$

所以 $x\in\mathscr{M}(P)$，这就证明了 $\mathscr{M}(P_1)\subset\mathscr{M}(P)$. 同理可证 $\mathscr{M}(P_2)\subset\mathscr{M}(P)$. 于是

$$\mathscr{M}(P_1)+\mathscr{M}(P_2)\subset\mathscr{M}(P)=\mathscr{M}(P_1+P_2).$$

但另一方面，任给 $x\in\mathscr{M}(P_1)\cap\mathscr{M}(P_2)$，则 $x=P_1x=P_2x=P_2P_1x=0$，于是 $\mathscr{M}(P_1)\cap\mathscr{M}(P_2)=\{0\}$. 再结合 $r(P_1)+r(P_2)\geqslant r(P_1+P_2)$，(2.6.15)得证.

设 $x\in\mathscr{N}(P_1+P_2)=\mathscr{N}(P)$，则 $Px=0$. 于是

$$P_1x = P_1^2x = P_1Px = 0,$$

这表明 $x\in\mathscr{N}(P_1)$. 类似地可证 $x\in\mathscr{N}(P_2)$. 这就证明了

$$\mathscr{N}(P_1+P_2)\subset\mathscr{N}(P_1)\cap\mathscr{N}(P_2). \qquad (2.6.17)$$

反过来，若 $x\in\mathscr{N}(P_1)\cap\mathscr{N}(P_2)$，则

$$Px = P_1x + P_2x = 0,$$

于是 $x\in\mathscr{N}(P)=\mathscr{N}(P_1+P_2)$，因而

$$\mathscr{N}(P_1)\cap\mathscr{N}(P_2)\subset\mathscr{N}(P_1+P_2),$$

结合(2.6.17)就证明了(2.6.16). 定理证毕.

定理 2.6.5 (a)设 P_1 和 P_2 为两个投影阵，则 $P=P_1-P_2$ 为投影阵，当且仅当

$$P_1P_2 = P_2P_1 = P_2. \qquad (2.6.18)$$

(b)若 $P_1=P_{S_1,T_1}$，$P_2=P_{S_2,T_2}$，且(2.6.18)成立，则

$$P \stackrel{\mathrm{d}}{=} P_1 - P_2 = P_{S,T}, \qquad (2.6.19)$$

其中

$$S = S_1\cap T_2,\quad T = T_1\oplus S_2.$$

证明 (a) 充分性　若 (2.6.18) 成立，则

$$P^2 = (P_1-P_2)^2 = P_1^2 - P_1P_2 - P_2P_1 + P_2^2$$

$$= \boldsymbol{P}_1 - \boldsymbol{P}_2 = \boldsymbol{P},$$

故 $\boldsymbol{P}$ 为投影阵.

必要性　若 $\boldsymbol{P}=\boldsymbol{P}_1-\boldsymbol{P}_2$ 为投影阵，将 $\boldsymbol{I}-\boldsymbol{P}$ 改写为

$$\boldsymbol{I} - \boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{P}_1) + \boldsymbol{P}_2, \tag{2.6.20}$$

这是两个投影阵 $\boldsymbol{I}-\boldsymbol{P}_1$ 与 $\boldsymbol{P}_2$ 之和. 根据定理 2.6.4，$\boldsymbol{I}-\boldsymbol{P}$ 为投影阵，必有

$$(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{P}_2 = \boldsymbol{P}_2(\boldsymbol{I} - \boldsymbol{P}_1) = \boldsymbol{0},$$

这就导致了 (2.6.18).

(b) 从 $\boldsymbol{I}-\boldsymbol{P}$ 的分解式 (2.6.20) 及定理 2.6.4 得

$$\begin{aligned}\boldsymbol{I} - \boldsymbol{P} &= (\boldsymbol{I} - \boldsymbol{P}_1) + \boldsymbol{P}_2 \\ &= P_{\mathscr{M}(\boldsymbol{I}-\boldsymbol{P}_1)\oplus\mathscr{M}(\boldsymbol{P}_2),\ \mathscr{N}(\boldsymbol{I}-\boldsymbol{P}_1)\cap\mathscr{N}(\boldsymbol{P}_2)}.\end{aligned}$$

于是

$$\boldsymbol{P} = \boldsymbol{P}_{\mathscr{N}(\boldsymbol{I}-\boldsymbol{P}_1)\cap\mathscr{N}(\boldsymbol{P}_2),\ \mathscr{M}(\boldsymbol{I}-\boldsymbol{P}_1)\oplus\mathscr{M}(\boldsymbol{P}_2)} \tag{2.6.21}$$

利用定理 2.6.1 (d) 和 (e)，有

$$\mathscr{M}(\boldsymbol{I} - \boldsymbol{P}_1) = \mathscr{N}(\boldsymbol{P}_1) = T_1,$$

$$\mathscr{N}(\boldsymbol{I} - \boldsymbol{P}_1) = \mathscr{M}(\boldsymbol{P}_1) = S_1.$$

此外，$\mathscr{N}(\boldsymbol{P}_2)=T_2$，$\mathscr{M}(\boldsymbol{P}_2)=S_2$. 将这些关系代入(2.6.21)即得(2.6.19). 定理证毕.

定理 2.6.6　(a) 设 $\boldsymbol{P}_1$ 和 $\boldsymbol{P}_2$ 为两个投影阵，则 $\boldsymbol{P} \stackrel{\mathrm{d}}{=} \boldsymbol{P}_1\boldsymbol{P}_2$ 为投影阵，当且仅当

$$\boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_2\boldsymbol{P}_1. \tag{2.6.22}$$

(b) 若 $\boldsymbol{P}_1=\boldsymbol{P}_{S_1,T_1}$，$\boldsymbol{P}_2=\boldsymbol{P}_{S_2,T_2}$，且 (2.6.22) 成立，则

$$\boldsymbol{P} \stackrel{\mathrm{d}}{=} \boldsymbol{P}_1\boldsymbol{P}_2 = \boldsymbol{P}_{S,T},$$

其中

$$T = T_1 + T_2, \quad S = S_1 \cap S_2.$$

证明 (a) 的证明很容易，留给读者作练习.

(b) 若(2.6.22)成立，则 $\boldsymbol{P}$ 为投影阵，由推论 2.6.1 知，$S=\mathscr{M}(\boldsymbol{P})$，$T=\mathscr{N}(\boldsymbol{P})$. 于是问题归结为证明

$$\mathscr{M}(\boldsymbol{P})=S_1\cap S_2, \tag{2.6.23}$$

$$\mathscr{N}(\boldsymbol{P})=T_1+T_2, \tag{2.6.24}$$

任给 $\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P})=\mathscr{M}(\boldsymbol{P}_1\boldsymbol{P}_2)$，则 $\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x}$. 于是

$$\boldsymbol{P}_1\boldsymbol{x}=\boldsymbol{P}_1^2\boldsymbol{P}_2\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x}=\boldsymbol{x},$$

这表明 $\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P}_1)=S_1$. 同理可证，$\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P}_2)=S_2$. 这就证明了

$$\mathscr{M}(\boldsymbol{P})\subset S_1\cap S_2. \tag{2.6.25}$$

反过来，若 $\boldsymbol{x}\in S_1\cap S_2$，即 $\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P}_1)\cap\mathscr{M}(\boldsymbol{P}_2)$，则 $\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{x}=\boldsymbol{P}_2\boldsymbol{x}$. 于是

$$\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x},$$

即 $\boldsymbol{x}\in\mathscr{M}(\boldsymbol{P}_1\boldsymbol{P}_2)=\mathscr{M}(\boldsymbol{P})$. 这就证明了$(S_1\cap S_2)\subset\mathscr{M}(\boldsymbol{P})$. 结合(2.6.25)，我们证明了(2.6.23).

下证(2.6.24). 任给 $\boldsymbol{x}\in\mathscr{N}(\boldsymbol{P})=\mathscr{N}(\boldsymbol{P}_1\boldsymbol{P}_2)$，则 $\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x}=\boldsymbol{0}$，这表明 $\boldsymbol{P}_2\boldsymbol{x}\in\mathscr{N}(\boldsymbol{P}_1)=T_1$. 此外，显然有

$$(\boldsymbol{I}-\boldsymbol{P}_2)\boldsymbol{x}\in\mathscr{N}(\boldsymbol{P}_2)=T_2,$$

故从分解式

$$\boldsymbol{x}=\boldsymbol{P}_2\boldsymbol{x}+(\boldsymbol{I}-\boldsymbol{P}_2)\boldsymbol{x},$$

知 $\boldsymbol{x}\in T_1+T_2$. 这就证明了

$$\mathscr{N}(\boldsymbol{P})\subset T_1+T_2. \tag{2.6.26}$$

另一方面，若 $\boldsymbol{x}\in T_1+T_2$，则 $\boldsymbol{x}$ 可分解为

$$\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2,$$

其中 $\boldsymbol{x}_i\in T_i$，$i=1,2$，则

$$\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x}=\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{x}_1=\boldsymbol{P}_2\boldsymbol{P}_1\boldsymbol{x}_1=\boldsymbol{0},$$

于是 $\boldsymbol{x}\in\mathscr{N}(\boldsymbol{P}_1\boldsymbol{P}_2)=\mathscr{N}(\boldsymbol{P})$，这就证明了 $T_1+T_2\subset\mathscr{N}(\boldsymbol{P})$. 结

合(2.6.26),就证明了(2.6.24).定理证毕.

前面我们讨论的都是一般投影阵,然而,在投影阵中有一类很重要的特殊的投影阵——正交投影阵.

若在(2.6.1)中,$S_1 \perp S_2$,即 S_1 和 S_2 互为正交补空间,则相应的投影 $\boldsymbol{y}$ 称为 $\boldsymbol{x}$ 向 S_1 的正交投影,相应的投影阵称为正交投影阵,简记为 $\boldsymbol{P}_{S_1}$.关于正交投影阵,我们有如下重要结论.

定理 2.6.7 设 C^n 中内积定义为

$$(\boldsymbol{x},\boldsymbol{y})_M = \boldsymbol{y}^* M \boldsymbol{x},$$

其中 $M>0$,则 $\boldsymbol{P}$ 为正交投影阵,当且仅当

(a) $\boldsymbol{P}^2=\boldsymbol{P}$;

(b) $M\boldsymbol{P}$ 为 Hermite 阵.

证明 必要性 从定理 2.6.2 知,当 $\boldsymbol{P}$ 为正交投影阵时,(a)成立.假设 $\boldsymbol{P}$ 为向子空间 S 的正交投影阵,则对任给的 $\boldsymbol{x}\in C^n$,$\boldsymbol{y}\in C^n$,有

$$\boldsymbol{P}\boldsymbol{x} \in S, \quad (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y} \in S^{\perp}.$$

于是

$$\boldsymbol{x}^* \boldsymbol{P}^* M(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y} = \boldsymbol{0},$$

对一切 $\boldsymbol{x},\boldsymbol{y}\in C^n$,这表明

$$\boldsymbol{P}^* M(\boldsymbol{I}-\boldsymbol{P}) = \boldsymbol{0},$$

上式等价于 $\boldsymbol{P}^* M=\boldsymbol{P}^* M\boldsymbol{P}$.因此式右端是 Hermite 阵,所以 $\boldsymbol{P}^* M=M\boldsymbol{P}$.(b)得证.

充分性 从(a)和推论 2.6.1 以及定理 2.6.1,得

$$\boldsymbol{P} = \boldsymbol{P}_{\mathcal{M}(\boldsymbol{P}),\mathcal{N}(\boldsymbol{P})} = \boldsymbol{P}_{\mathcal{M}(\boldsymbol{P}),\mathcal{M}(\boldsymbol{I}-\boldsymbol{P})}.$$

问题归结为证明

$$\mathcal{M}(\boldsymbol{I}-\boldsymbol{P}) = \mathcal{N}(\boldsymbol{P})^{\perp} \tag{2.6.27}$$

事实上,因为 $M\boldsymbol{P}$ 为 Hermite 阵,故有 $M\boldsymbol{P}=\boldsymbol{P}^* M$,用 $\boldsymbol{P}^*$ 左

乘，并利用 P^* 为幂等阵可得

$$P^*MP = P^*M. \tag{2.6.28}$$

对任给的 $x\in\mathscr{M}(I-P)$，必存在 $u\in C^n$，使得 $x=(I-P)u$. 利用(2.6.28)，我们有

$$t^*P^*M(I-P)u = 0,$$

对一切 $t\in C^n$ 成立. 这就证明了 $x=(I-P)u\in\mathscr{M}(P)^{\perp}$，从而证明了

$$\mathscr{M}(I-P) \subset \mathscr{M}(P)^{\perp}. \tag{2.6.29}$$

再应用定理 2.6.1(c)，有

$$\begin{aligned}\dim\mathscr{M}(I-P) &= r(I-P) = \mathrm{tr}(I-P) \\ &= n - \mathrm{tr}P = n - r(P) \\ &= \dim\mathscr{M}(P)^{\perp}.\end{aligned}$$

结合(2.6.29)就得到(2.6.27). 定理证毕.

推论 2.6.2 若在 C^n 中定义标准内积 $(x,y)=y^*x$，则 P 为正交投影阵当且仅当 P 为幂等 Hermite 阵，即 P 满足

(a) $P^*=P$;

(b) $P^2=P$.

应用广义逆矩阵，可以给出投影阵和正交投影阵的具体表示，这使得我们能够更方便地研究投影阵及正交投影阵的性质. 这些将在 § 3.4 中讨论.

§ 2.7 谱 分 解

如果一个方阵相似于一个对角阵，则称该方阵是可对角化的. 本节的目的是证明这样的方阵可表为若干个投影阵的线性组合. 特别，对正规阵，它可表为正交投影阵的线性组合.

定理 2.7.1(谱分解) 设 $A\in C^{n\times n}$，$\lambda_1,\cdots,\lambda_k$ 为 A 的全部

不同特征值，则 $\boldsymbol{A}$ 是可对角化的，当且仅当存在投影阵(即幂等阵)$\boldsymbol{P}_1,\cdots,\boldsymbol{P}_k$，使得

$$\boldsymbol{P}_i\boldsymbol{P}_j=\boldsymbol{0},\qquad i\neq j, \tag{2.7.1}$$

$$\boldsymbol{I}_n=\sum_{i=1}^{k}\boldsymbol{P}_i, \tag{2.7.2}$$

$$\boldsymbol{A}=\sum_{i=1}^{k}\lambda_i\boldsymbol{P}_i. \tag{2.7.3}$$

证明 充分性 设 $r(\boldsymbol{P}_i)=r_i, i=1,\cdots,k$. 又设 $\boldsymbol{X}_i$ 为 $n\times r_i$ 的矩阵，满足 $\mathscr{M}(\boldsymbol{X}_i)=\mathscr{M}(\boldsymbol{P}_i)$. 记

$$\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_k).$$

因为 $\boldsymbol{P}_i$ 为幂等阵，应用定理 2.6.1(c)，则矩阵 $\boldsymbol{X}$ 的列数为

$$\sum_{i=1}^{k}r_i=\sum_{i=1}^{k}r(\boldsymbol{X}_i)=\sum_{i=1}^{k}r(\boldsymbol{P}_i)=\sum_{i=1}^{k}\mathrm{tr}(\boldsymbol{P}_i)=\mathrm{tr}(\sum_{i=1}^{k}\boldsymbol{P}_i)$$
$$=\mathrm{tr}\boldsymbol{I}_n=n.$$

于是 $\boldsymbol{X}$ 为 n 阶方阵.

依 $\boldsymbol{X}_i$ 的定义，对每个 $\boldsymbol{X}_i$，均存在 $\boldsymbol{U}_i$，使得 $\boldsymbol{P}_i=\boldsymbol{X}_i\boldsymbol{U}_i, i=1,\cdots,k$. 定义

$$\boldsymbol{U}'=(\boldsymbol{U}'_1,\boldsymbol{U}'_2,\cdots,\boldsymbol{U}'_k),$$

利用(2.7.2)，有

$$\boldsymbol{X}\boldsymbol{U}=\sum_{i=1}^{k}\boldsymbol{X}_i\boldsymbol{U}_i=\sum_{i=1}^{k}\boldsymbol{P}_i=\boldsymbol{I}_n,$$

这表明 $\boldsymbol{X}$ 是可逆阵. 应用(2.7.1)和(2.7.3)，我们有

$$\boldsymbol{A}\boldsymbol{X}=(\sum\lambda_i\boldsymbol{P}_i)\boldsymbol{X}=(\lambda_1\boldsymbol{X}_1,\cdots,\lambda_k\boldsymbol{X}_k)=\boldsymbol{X}\boldsymbol{D}, \tag{2.7.4}$$

这里利用了 $\boldsymbol{P}_i\boldsymbol{X}_i=\boldsymbol{X}_i, \boldsymbol{P}_i\boldsymbol{X}_j=\boldsymbol{0}, j\neq i$，且

$$\boldsymbol{D}=\mathrm{diag}(\lambda_1\boldsymbol{I}_{r_1},\lambda_2\boldsymbol{I}_{r_2},\cdots,\lambda_k\boldsymbol{I}_{r_k}). \tag{2.7.5}$$

因为已证 $\boldsymbol{X}$ 为可逆阵，于是从(2.7.4)可推出

$$\boldsymbol{A}=\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{-1},$$

即 $\boldsymbol{A}$ 相似于对角阵 $\boldsymbol{D}$.

必要性 设 $\boldsymbol{A}$ 是可对角化的，且设 r_i 为特征值 λ_i 的重数，$i=1,\cdots,k$. 则存在可逆阵 $\boldsymbol{X}$，使得

$$\boldsymbol{AX}=\boldsymbol{XD},$$

这里 $\boldsymbol{D}$ 由(2.7.5)定义. 将矩阵 $\boldsymbol{X}$ 分块：

$$\boldsymbol{X}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_k),$$

其中 $\boldsymbol{X}_i$ 为 $n\times r_i$ 矩阵. 定义

$$\begin{aligned}\boldsymbol{P}_i&=(\boldsymbol{0},\cdots,\boldsymbol{0},\boldsymbol{X}_i,\boldsymbol{0},\cdots,\boldsymbol{0})\boldsymbol{X}^{-1}\\&\stackrel{\mathrm{d}}{=}\boldsymbol{Y}_i\boldsymbol{X}^{-1},\quad i=1,\cdots,k.\end{aligned}$$

则 $\boldsymbol{P}_i$ 为幂等阵. 事实上，$\boldsymbol{Y}_i$ 可改写为

$$\boldsymbol{Y}_i=\boldsymbol{X}\mathrm{diag}(\boldsymbol{0},\cdots,\boldsymbol{0},\boldsymbol{I}_{r_i},\boldsymbol{0},\cdots,\boldsymbol{0}),$$

据此易证 $\boldsymbol{P}_i$ 为幂等阵且满足(2.7.1)—(2.7.3). 定理证毕.

定理 2.7.2 设 $\boldsymbol{A}\in\boldsymbol{C}^{n\times n}$是可对角化方阵，$\lambda_1,\cdots,\lambda_k$ 为 $\boldsymbol{A}$ 的全部互不相同的特征值. 则满足(2.7.1)—(2.7.3)的幂等阵 $\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_k$ 由 $\boldsymbol{A}$ 唯一确定.

证明 设 $\boldsymbol{Q}_i,i=1,\cdots,k$ 为幂等阵，且满足

$$\boldsymbol{Q}_i\boldsymbol{Q}_j=\boldsymbol{0},\quad i\neq j,\tag{2.7.6}$$

$$\boldsymbol{I}_n=\sum_{i=1}^{k}\boldsymbol{Q}_i,\tag{2.7.7}$$

$$\boldsymbol{A}=\sum_{i=1}^{k}\lambda_i\boldsymbol{Q}_i.\tag{2.7.8}$$

我们将证明：$\boldsymbol{P}_i=\boldsymbol{Q}_i,i=1,\cdots,k$.

从(2.7.1)和(2.7.3)，可得

$$\boldsymbol{P}_i\boldsymbol{A}=\boldsymbol{AP}_i=\lambda_i\boldsymbol{P}_i,\quad i=1,\cdots,k,\tag{2.7.9}$$

同样，由(2.7.6)和(2.7.7)得到

$$\boldsymbol{Q}_i\boldsymbol{A}=\boldsymbol{AQ}_i=\lambda_i\boldsymbol{Q}_i,\quad i=1,\cdots,k,\tag{2.7.10}$$

于是

$$P_i(AQ_j)=\lambda_j P_i Q_j,$$

$$(P_iA)Q_j=\lambda_i P_i Q_j.$$

因为当 $i\neq j$ 时，$\lambda_i\neq\lambda_j$，上两式证明了

$$P_iQ_j=0,\quad i\neq j.$$

从而

$$P_i=P_i\left(\sum_{j=1}^{k}Q_j\right)=P_iQ_i=\left(\sum_{j=1}^{k}P_j\right)Q_i=Q_i.$$

定理证毕.

注 前面两定理中的幂等阵 P_i 称为 A 的主幂等阵(principal idempotent matrix). $\mathscr{M}(P_i)$称为 A 的特征子空间.

下面的定理给出了一种求主幂等阵的方法.

定理 2.7.3 设 A 为 n 阶可对角化方阵，$\lambda_1,\cdots,\lambda_k$ 为它的全部不同特征值. 记

$$\varphi_i(\lambda)=\prod_{\substack{t=1\\t\neq i}}^{k}(\lambda-\lambda_t).$$

则

$$G_i=\frac{\varphi_i(A)}{\varphi_i(\lambda_i)},\quad i=1,\cdots,k$$

为 A 的主幂等阵，这里 $\varphi_i(A)$定义为

$$\varphi_i(A)=\prod_{\substack{t=1\\t\neq i}}^{k}(A-\lambda_t I_n).$$

证明 设 $P_1,\cdots,P_k$ 为 A 的主幂等阵. 则

$$G_iP_j=\frac{1}{\varphi_i(\lambda_i)}\prod_{\substack{t=1\\t\neq i}}^{k}(A-\lambda_t I_n)P_j$$

$$=\frac{1}{\varphi_i(\lambda_i)}\prod_{\substack{t=1\\t\neq i}}^{k}(\lambda_j-\lambda_t)P_j$$

$$= \begin{cases} P_j, & 若\ i = j, \\ 0, & 若\ i \neq j. \end{cases}$$

于是

$$G_i = G_i \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} G_i P_j = P_i, \quad i = 1, \cdots, k.$$

定理证毕.

推论 2.7.1 设 $A \in C^{n\times n}$是可对角化阵，它的全部不同的特征值为 $\lambda_1, \cdots, \lambda_k$，对应的主幂等阵为 $P_1, \cdots, P_k$. 则

(a)对任一多项式 $f(x)$，

$$f(A) = \sum_{i=1}^{k} f(\lambda_i) P_i.$$

(b)方阵 B 与 A 可交换，当且仅当 B 与每个 P_i 可交换.

结论容易从定理 2.7.1 和定理 2.7.3 导出. 请读者写出证明的细节.

我们知道，正规阵是一类特殊的可对角化方阵. 下面的定理给出正规阵的谱分解.

定理 2.7.4 假定在 C^n 中内积定义为$(x, y) = y^* x$. 设 $A \in C^{n\times n}$，$\lambda_1, \cdots, \lambda_k$ 为 A 的全部互不相同的特征值. 则 A 为正规阵，当且仅当存在正交投影阵 $P_1, \cdots, P_k$，使得

$$P_i P_j = 0, \quad i \neq j, \tag{2.7.11}$$

$$I = \sum_{i=1}^{k} P_i \tag{2.7.12}$$

$$A = \sum_{i=1}^{k} \lambda_i P_i. \tag{2.7.13}$$

证明 充分性 假定 A 有分解式(2.7.13)，其中 P_i 为正交投影阵，且满足(2.7.11)和(2.7.12). 因为内积是标准内积，依推论 2.6.2，P_i 为幂等 Hermite 阵. 于是

$$AA^* = \left(\sum_{i=1}^{k}\lambda_i P_i\right)\left(\sum_{i=1}^{k}\overline{\lambda}_i P_i\right) = \sum_{i=1}^{k}|\lambda_i|^2 P_i = A^* A.$$

必要性　设 A 为正规阵，则它是可对角化的. 依定理 2.7.1，存在投影阵 $P_1,\cdots,P_k$，满足(2.7.11)、(2.7.12)和(2.7.13). 我们只需证明：P_i 为 Hermite 阵.

因为 A 是正规阵. 于是 A 的对应于 λ_i 的特征子空间 $\mathcal{M}(P_i)$也是 A^* 的对应于特征值$\overline{\lambda}_i$ 的特征子空间. 因而 A^* 与 A 有相同的主幂等阵. 依定理 2.7.1，A^* 有分解式：

$$A^* = \sum_{i=1}^{k}\overline{\lambda}_i P_i. \tag{2.7.14}$$

在(2.7.13)的两边取转置共轭，得

$$A^* = \sum_{i=1}^{k}\overline{\lambda}_i P_i^*. \tag{2.7.15}$$

注意到，P_i^* 满足

$$P_i^* P_j^* = 0,\quad i \neq j,$$

$$I = \sum_{i=1}^{k} P_i^*.$$

这表明(2.7.15)也是 A^* 的一个谱分解. 由分解的唯一性、(2.7.14)和(2.7.15)知 $P_i = P_i^*$，$i=1,\cdots,k$，即 P_i 为 Hermite 阵，定理证毕.

§ 2.8　特征值的极值性质

设 A 为 $n\times n$ Hermite 阵. 对任给的 $x\in C^n$，我们称

$$\frac{x^* Ax}{x^* x} \tag{2.8.1}$$

为 Rayleigh-Ritz 商或 Rayleigh 商. 在矩阵论中有两个重要定理，即 Rayleigh-Ritz 定理和 Courant-Fischer 定理，它们通

过 Rayleigh 商(2.8.1)的无约束和约束的极值来表征 Hermite 阵的特征值. 这些结果不仅在许多领域具有广泛应用，而且在广义逆矩阵研究中也很有用处. 本节的目的就是简要讨论这两个重要定理，关于它们的一些应用可在王松桂等(1994)中找到.

定理 2.8.1(Rayleigh-Ritz)　设 $\boldsymbol{A}$ 为 n 阶 Hermite 阵，$\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 $\boldsymbol{A}$ 的特征值. 则

(a)　$\lambda_n \boldsymbol{x}^* \boldsymbol{x} \leqslant \boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x} \leqslant \lambda_1 \boldsymbol{x}^* \boldsymbol{x}$；

(b)　$\lambda_1 = \max\limits_{x \neq 0} \dfrac{\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^* \boldsymbol{x}} = \max\limits_{x^* x = 1} \boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}$；

(c)　$\lambda_n = \min\limits_{x \neq 0} \dfrac{\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^* \boldsymbol{x}} = \min\limits_{x^* x = 1} \boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}$.

证明　(a)　设 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ 为 $\boldsymbol{A}$ 的对应于 $\lambda_1, \cdots, \lambda_n$ 的标准正交化特征向量. 记 $\boldsymbol{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$，则 $\boldsymbol{U}$ 为一酉阵. 显然 $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^*$，其中 $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 记 $\boldsymbol{y} = \boldsymbol{U}^* \boldsymbol{x}$，则 $\boldsymbol{y}^* \boldsymbol{y} = \boldsymbol{x}^* \boldsymbol{x}$. 于是

$$\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^* \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^* \boldsymbol{x} = \boldsymbol{y}^* \boldsymbol{\Lambda} \boldsymbol{y} = \sum_{i=1}^{n} \lambda_i |y_i|^2.$$

注意到

$$\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x} \leqslant \lambda_1 \sum_{i=1}^{n} |y_i|^2 = \lambda_1 \boldsymbol{y}^* \boldsymbol{y} = \lambda_1 \boldsymbol{x}^* \boldsymbol{x},$$

$$\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x} \geqslant \lambda_n \sum_{i=1}^{n} |y_i|^2 = \lambda_n \boldsymbol{y}^* \boldsymbol{y} = \lambda_n \boldsymbol{x}^* \boldsymbol{x}.$$

当 $\boldsymbol{x} = \boldsymbol{u}_1$ 和 $\boldsymbol{x} = \boldsymbol{u}_n$ 时，上面两式等号成立，于是(a)得证. (b)和(c)是(a)的直接推论.

上面的定理刻画了 Hermite 阵最大和最小特征值的极值

性质.对于其它特征值,我们有

定理 2.8.2 设 $\boldsymbol{A}$ 为 $n\times n$ Hermite 阵,$\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 $\boldsymbol{A}$ 的特征值,$\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$ 为对应的标准正交化特征向量.则

(a) $$\lambda_k=\max_{\substack{x\neq 0\\ u_j^* x=0\\ j=1,\cdots,k-1}}\frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}=\max_{\substack{x^*x=1\\ u_j^* x=0\\ j=1,\cdots,k-1}}\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x};$$

(b) $$\lambda_{n-k}=\min_{\substack{x\neq 0\\ u_j^* x=0\\ j=n-k+1,\cdots,n}}\frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}=\max_{\substack{x^*x=1\\ u_j^* x=0\\ j=n-k+1,\cdots,n}}\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}.$$

证明 (a) 证明类似于定理 2.8.1.只要注意到,对现在的情况,$\boldsymbol{y}=\boldsymbol{U}^*\boldsymbol{x}=(y_1,\cdots,y_n)'$,$y_1=\cdots=y_{k-1}=0$.于是

$$\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}=\sum_{i=k}^{n}\lambda_i|y_i|^2.$$

(b) 证明类似于(a).定理证毕.

这两个定理虽然刻画了 Hermite 阵的特征值的极值性质,但是在后一个定理中,所求的是 $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}$ 或 $\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}/(\boldsymbol{x}^*\boldsymbol{x})$ 的条件极值,并且所用条件依赖于方阵 $\boldsymbol{A}$ 的特征向量,这在应用上很不方便.下面的 Courant-Fischer 的 min-max 定理克服了这个缺陷.

定理 2.8.3(Courant-Fischer) 设 $\boldsymbol{A}$ 为 $n\times n$ Hermite 阵,$\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 $\boldsymbol{A}$ 的特征值,$\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$ 为对应的标准正交化特征向量,则

(a) $$\min_{B}\max_{\substack{x\neq 0\\ B^*x=0}}\frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}=\lambda_k,$$

其中 $\boldsymbol{B}$ 为 $n\times(k-1)$矩阵.当 $\boldsymbol{B}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{k-1})$时,上式达到 λ_k.

(b) $$\min_{B}\max_{\substack{x\neq 0\\ B^*x=0}}\frac{\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}=\lambda_{n-k+1},$$

其中 B 为 $n\times(k-1)$ 矩阵. 当 $B=(u_{n-k+2},\cdots,u_n)$ 时,上式达到 λ_{n-k+1}.

证明(a) 记 $U=(u_1,\cdots,u_n)$, $y=U^*x$, $D=U^*B$, $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$. 于是

$$\max_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}=\max_{\substack{y\neq 0\\ D^*y=0}}\frac{y^*\Lambda y}{y^*y}=\max_{\substack{y^*y=1\\ D^*y=0}}\sum_{i=1}^{n}\lambda_i|y_i|^2$$

$$\geqslant\max_{\substack{y^*y=1\\ D^*y=0\\ y_{k+1}=\cdots=y_n=0}}\sum_{i=1}^{n}\lambda_i|y_i|^2=\max_{\substack{D^*y=0\\ |y_1|^2+\cdots+|y_k|^2=1}}\sum_{i=1}^{k}\lambda_i|y_i|^2$$

$$\geqslant\min_{\substack{D^*y=0\\ |y_1|^2+\cdots+|y_k|^2=1}}\sum_{i=1}^{k}\lambda_i|y_i|^2\geqslant\min_{|y_1|^2+\cdots+|y_k|^2=1}\sum_{i=1}^{k}\lambda_i|y_i|^2=\lambda_k.$$

即我们证明了,对一切 $n\times(k-1)$ 矩阵 B,

$$\max_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}\geqslant\lambda_k,$$

因此

$$\min_{B}\max_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}\geqslant\lambda_k.$$

结合定理 2.8.2(a),知当 $B=(u_1,\cdots,u_{k-1})$ 时,等号成立. (a) 得证.

(b) 采用与(a)相同的记号,则

$$\min_{\substack{x\neq 0\\ B^*x=0}}\frac{x^*Ax}{x^*x}=\min_{\substack{y\neq 0\\ D^*y=0}}\frac{y^*\Lambda y}{y^*y}=\min_{\substack{y^*y=1\\ D^*y=0}}\sum_{i=1}^{n}\lambda_i|y_i|^2$$

$$\leqslant\min_{\substack{D^*y=0\\ y^*y=1\\ y_1=\cdots=y_{n-k}=0}}\sum_{i=1}^{n}\lambda_i|y_i|^2=\min_{\substack{D^*y=0\\ |y_{n-k+1}|^2+\cdots+|y_n|^2=1}}\sum_{i=n-k+1}^{n}\lambda_i|y_i|^2$$

$$\leqslant \max_{|y_{n-k+1}|^2+\cdots+|y_n|^2=1} \sum_{i=n-k+1}^{n} \lambda_i |y_i|^2 = \lambda_{n-k+1}.$$

至此，我们证明了，对一切 $n\times(k-1)$ 矩阵 $\boldsymbol{B}$，

$$\min_{\substack{x\neq 0\\ B^*x=0}} \frac{x^*Ax}{x^*x} \leqslant \lambda_{n-k+1},$$

因此

$$\max_{B} \min_{\substack{x\neq 0\\ B^*x=0}} \frac{x^*Ax}{x^*x} \leqslant \lambda_{n-k+1}.$$

结合定理 2.8.2(b)，上式当 $\boldsymbol{B}=(u_{n-k+2},\cdots,u_n)$ 时等号成立. 定理证毕.

§2.9 矩阵的范数

设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，所谓 $\boldsymbol{A}$ 的范数，记为 $\|\boldsymbol{A}\|$，就是满足下列三条性质的任一非负函数：

(a) 正定性：$\|\boldsymbol{A}\|\geqslant 0$，且 $\|\boldsymbol{A}\|=0\Leftrightarrow \boldsymbol{A}=\boldsymbol{0}$；

(b) 齐次性：$\|c\boldsymbol{A}\|=|c|\,\|\boldsymbol{A}\|$，其中 c 为任一复数；

(c) 三角不等式：$\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$.

特别，若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别为 $m\times n$ 和 $n\times p$ 矩阵，$\|\boldsymbol{A}\|$ 满足

$$\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|,$$

则称该范数是相容的.

在后面的讨论中，下面的两种范数要常常用到.

例 2.9.1 欧氏范数，也称 Frobenius 范数，记为

$$\|\boldsymbol{A}\|_F = (\mathrm{tr}\boldsymbol{A}^*\boldsymbol{A})^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2},$$

这里 $\boldsymbol{A}\in \boldsymbol{C}^{m\times n}$. 不难看出，欧氏范数就是把矩阵的列一个接一个排起来，组成 $mn\times 1$ 的列向量，然后求该列向量的欧氏长

度.

欧氏范数是相容范数,此因

$$\begin{aligned}\|AB\|_F^2 &= \mathrm{tr}(B^*A^*AB) \leqslant \mathrm{tr}(\lambda_1(A^*A)B^*B)\\ &= \lambda_1(A^*A)\mathrm{tr}(B^*B) \leqslant \mathrm{tr}(A^*A)\mathrm{tr}(B^*B)\\ &= \|A\|_F^2\|B\|_F^2,\end{aligned}$$

其中 $\lambda_1(C)$ 表示矩阵 C 的最大特征值.

以后我们也常把 $\|A\|_F$ 简记为 $\|A\|$,读者从上下文是不会引起混淆的.

例 2.9.2 谱范数

$$\|A\|_2 = \lambda_1^{1/2}(A^*A) = \lambda_1^{1/2}(A^*A),$$

这里 $\lambda_1(B)$ 表示矩阵 B 的最大特征值.

在定理 2.4.2 中我们定义了矩阵 A 的奇异值. 易见矩阵 A 的谱范数在数值上等于 A 的最大奇异值. 利用方阵特征值的单调性:若 $A\geqslant 0$,$B\geqslant 0$,且 $A\geqslant B$(即 $A-B\geqslant 0$),则 $\lambda_1(A)\geqslant\lambda_1(B)$(证明见王松桂等(1994). p. 114),容易证明谱范数也是相容范数. 事实上,

$$\begin{aligned}\|AB\|_2^2 &= \lambda_1(B^*A^*AB) \leqslant \lambda_1(B^*\lambda_1(A^*A)B)\\ &\leqslant \lambda_1(A^*A)\lambda_1(B^*B)\\ &= \|A\|_2^2\|B\|_2^2.\end{aligned}$$

还有一类矩阵范数,称之为加权范数,在广义逆研究中很有用.

设 $A\in C^{m\times n}$,它是从 C^n 到 C^m 的一个映射. 若 $x\in C^n$,则 $y=Ax\in C^m$,若用 $(x,y)_M$ 和 $(x,y)_N$ 分别表示 C^m 和 C^n 的内积,即

$$(x,y)_M = y^*Mx, \qquad \forall\, x,y\in C^m,$$

$$(x,y)_N = y^*Nx, \qquad \forall\, x,y\in C^n,$$

其中 M 和 N 分别为 $m\times m, n\times n$ 的 Hermite 正定阵. 与这样的内积相对应的向量范数分别为

$$\|x\|_M=(x,x)_M^{1/2}=(x^*Mx)^{1/2}$$

$$=\|M^{1/2}x\|,\qquad \forall x\in C^m,$$

$$\|x\|_N=(x,x)_N^{1/2}=(x^*Nx)^{1/2}$$

$$=\|N^{1/2}x\|,\qquad \forall x\in C^n.$$

与上述两种向量范数相对应，我们可以定义 $m\times n$ 矩阵 A 的范数

$$\|A\|_{MN}=\max_{\|x\|_N=1}\|Ax\|_M,\qquad (2.9.1)$$

称为 A 的加权范数. 当 $M=I_m, N=I_n$, $\|A\|_{II}=\|A\|_2$. 利用定理 2.8.1，我们有

$$\|A\|_{MN}=\left(\max_{x\neq 0}\frac{x^*A^*MAx}{x^*Nx}\right)^{1/2}=\lambda_1^{1/2}(A^*MAN^{-1})$$

$$=\|M^{\frac{1}{2}}AN^{-\frac{1}{2}}\|_2\qquad (2.9.2).$$

加权范数也具有相容性. 它的表述略复杂些，可以写成如下定理.

定理 2.9.1 设 $A\in C^{m\times n}, B\in C^{n\times m}, x\in C^n, y\in C^m$. 则

(a) $\|Ax\|_M\leqslant\|A\|_{MN}\|x\|_N$,

(b) $\|By\|_N\leqslant\|B\|_{NM}\|y\|_M$,

(c) $\|AB\|_{MM}\leqslant\|A\|_{MN}\|B\|_{NM}$.

这些事实的证明是容易的，留作练习.

与 C^m 和 C^n 中定义的内积紧密联系的一个概念是加权共轭转置阵或称伴随阵(adjoint matrix).

定义 2.9.1 设 $A\in C^{m\times n}$, M 和 N 分别为 $m\times m$、$n\times n$

Hermite 正定阵. A 的加权共轭转置阵，记为 $A^{\#}$，定义为

$$(Ax,y)_M = (x,A^{\#}y)_N,$$

对任给 $x\in C^n$，$y\in C^m$. 不难证明

$$A^{\#} = N^{-1}A^*M. \tag{2.9.3}$$

当 $M=I_m$，$N=I_n$ 时，$A^{\#}=A^*$，即对这一特殊情况，$A^{\#}$ 就变为通常的转置共轭阵.

类似于 A^*，$A^{\#}$ 具有下列性质：

定理 2.9.2

(a) 设 A、$B\in C^{m\times n}$，则 $(A+B)^{\#}=A^{\#}+B^{\#}$.

(b) 设 $A\in C^{m\times n}$，$B\in C^{n\times m}$，则 $(AB)^{\#}=B^{\#}A^{\#}$.

(c) $(A^{\#})^{\#}=A$.

(d) $(A^{\#})^{-1}=(A^{-1})^{\#}$.

(e) 设 $A\in C^{m\times n}$，则 $\|A\|_{MN}=\|A^{\#}\|_{NM}$.

(f) 设 $A\in C^{m\times n}$，则

$$\|A\|_{MN}^2 = \|AA^{\#}\|_{MM} = \|A^{\#}A\|_{NN}.$$

这些事实的证明可根据 $A^{\#}$ 的定义或(2.9.1)和(2.9.2)来完成. 下面我们只给出(f)的证明.

由(2.9.2)得

$$\begin{aligned}\|AA^{\#}\|_{MM} &= \|M^{\frac{1}{2}}AA^{\#}M^{-\frac{1}{2}}\|_2 \\ &= \|M^{\frac{1}{2}}AN^{-\frac{1}{2}}(M^{\frac{1}{2}}AN^{-\frac{1}{2}})^*\|_2 \\ &= \|M^{\frac{1}{2}}AN^{-\frac{1}{2}}\|_2^2 \\ &= \|A\|_{MN}^2,\end{aligned}$$

这就证明了第一个等式. 用完全类似的方法可以证明第二个等式.

在本书中，加权范数主要用在几种重要的加权广义逆矩阵中，如加权 Moore-Penrose 广义逆(见 §4.9)、加权{1,3}-

逆和加权{1,4}-逆(见第五章).

§2.10 奇 异 值

在§2.4,我们引进了矩阵的奇异值. 设 A 为 $m\times n$ 矩阵,其秩为 r. 那么 A 的奇异值定义为 A^*A,或等价地,AA^* 的非零特征值的正平方根,本节讨论有关奇异值的几个重要事实,以备后续章节引用.

定理 2.10.1 设 A 为 $m\times n$ 阵,秩为 r,记 $\lambda_1\geqslant\cdots\geqslant\lambda_r>0$ 为 A 的奇异值. 设 r 个复数 d_i 满足

$$|d_i| = \lambda_i, \quad i = 1,\cdots,r.$$

设 $u_1,\cdots,u_r$ 为 AA^* 对应于非零特征值 $\lambda_i^2,i=1,\cdots,r$ 的标准正交化特征向量. 定义

$$\varphi_i = \frac{1}{\overline{d}_i}A^*u_i, \qquad i = 1,\cdots,r.$$

则

(a) $\varphi_1,\cdots,\varphi_r$ 为 A^*A 对应于非零特征值 $\lambda_i^2,i=1,\cdots,r$ 的标准正交化特征向量.

(b) $u_i=\frac{1}{d_i}A\varphi_i,i=1,\cdots,r.$

证明 (a) 由 φ_i 之定义,有

$$\begin{aligned}A^*A\varphi_i &= \frac{1}{\overline{d}_i}A^*(AA^*u_i) = \frac{1}{\overline{d}_i}A^*\lambda_i^2u_i \\ &= \frac{|d_i|^2}{\overline{d}_i}A^*u_i = |d_i|^2\varphi_i = \lambda_i^2\varphi_i.\end{aligned}$$

另一方面

$$\varphi_i^*\varphi_j = \frac{1}{d_i\overline{d}_j}u_i^*(AA^*u_j) = \frac{1}{d_i\overline{d}_j}u_i^*(\lambda_j^2u_j)$$

$$= \frac{\lambda_j^2}{d_i \overline{d}_j} u_i^* u_j = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

这就完成了(a)的证明.

(b)的证明是容易的,定理证毕.

注 本定理的对偶结论也成立. 即若 $\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_r$ 为 $\boldsymbol{A}^*\boldsymbol{A}$ 的对应于非零特征值 $\lambda_i^2, i=1,\cdots,r$ 的标准正交化特征向量,定义

$$\boldsymbol{u}_i = \frac{1}{d_i} \boldsymbol{A} \boldsymbol{\varphi}_i, \qquad i = 1, \cdots, r.$$

则 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r$ 为 $\boldsymbol{A}\boldsymbol{A}^*$ 对应于非零特征值 $\lambda_i^2, i=1,\cdots,r$ 的标准正交化特征向量,且

$$\boldsymbol{\varphi}_i = \frac{1}{\overline{d}_i} \boldsymbol{A}^* \boldsymbol{u}_i, \qquad i = 1, \cdots, r.$$

事实上,将定理中 $\boldsymbol{A}$ 与 $\boldsymbol{A}^*$ 互换,即得此对偶结论.

定理 2.10.2 设 $\boldsymbol{A}$ 为 $m \times n$ 阵,秩为 r, $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0$ 为 $\boldsymbol{A}$ 的奇异值,设 r 个复数 d_i 满足

$$|d_i| = \lambda_i, \quad i = 1, \cdots, r.$$

则存在 $m \times m$ 和 $n \times n$ 酉阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$,使得

$$\boldsymbol{P}^* \boldsymbol{A} \boldsymbol{Q} = \mathrm{diag}(d_1, \cdots, d_r, 0, \cdots, 0) \qquad (2.10.1)$$

证明 设 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r$ 为 $\boldsymbol{A}\boldsymbol{A}^*$ 的对应于非零特征值 $\lambda_i^2, i=1, \cdots, r$ 的标准正交化特征向量. 它们构成子空间 $\mathscr{M}(\boldsymbol{A}\boldsymbol{A}^*) = \mathscr{M}(\boldsymbol{A})$ 的一组基. 设 $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$ 为 $\mathscr{M}(\boldsymbol{A})^{\perp} = \mathscr{N}(\boldsymbol{A}^*)$ 的一组标准正交基. 于是向量组 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r, \boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$ 为 $\boldsymbol{C}^m$ 的一组标准正交基. 定义

$$\boldsymbol{P} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r, \boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m),$$

这是一个 $m \times m$ 酉阵,且

$$\boldsymbol{A}^* \boldsymbol{u}_i = \boldsymbol{0}, \quad i = r+1, \cdots, m. \qquad (2.10.2)$$

用完全类似的方法，在上面的构造过程中将 $\boldsymbol{A}$ 和 $\boldsymbol{A}^*$ 互换，可构造出标准正交化向量组 $\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_n$，满足

$$\boldsymbol{A}\boldsymbol{\varphi}_j = \boldsymbol{0}, \quad j = r+1,\cdots,n, \qquad (2.10.3)$$

定义

$$\boldsymbol{Q} = (\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_r,\boldsymbol{\varphi}_{r+1},\cdots,\boldsymbol{\varphi}_n),$$

这是一个 $n\times n$ 酉阵. 记

$$\boldsymbol{D} = (d_{ij}) = \boldsymbol{P}^*\boldsymbol{A}\boldsymbol{Q}.$$

我们证明 $\boldsymbol{D}$ 具有形式(2.10.1).

事实上，利用(2.10.2)和(2.10.3)，有

$$d_{ij} = \boldsymbol{u}_i^*\boldsymbol{A}\boldsymbol{\varphi}_j = 0,$$

对 $i>r$，或 $j>r$. 而当 $i,j=1,\cdots,r$ 时，利用上一定理中 $\boldsymbol{u}_i$ 与 $\boldsymbol{\varphi}_i$ 的关系

$$\begin{aligned} d_{ij} &= \boldsymbol{u}_i^*\boldsymbol{A}\boldsymbol{\varphi}_j = \frac{1}{d_j}\boldsymbol{u}_i^*(\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{u}_j) \\ &= \frac{1}{d_j}\boldsymbol{u}_i^*(\lambda_j^2\boldsymbol{u}_j) = d_j\boldsymbol{u}_i^*\boldsymbol{u}_j \\ &= \begin{cases} 0, & i\neq j, \\ d_j, & i=j. \end{cases} \end{aligned}$$

定理证毕.

最后，我们证明关于奇异值扰动的一个事实.

假设 $\boldsymbol{A}\in \boldsymbol{C}^{m\times n}$，其秩为 r. 在前面的讨论中 $\boldsymbol{A}$ 的奇异值定义为 $\boldsymbol{A}^*\boldsymbol{A}$，或等价地，$\boldsymbol{A}\boldsymbol{A}^*$ 的 r 个非零特征值的正平方根. 但为方便计，有时候人们也把 $\boldsymbol{A}^*\boldsymbol{A}$ 或 $\boldsymbol{A}\boldsymbol{A}^*$ 的全部特征值的非负平方根称为 $\boldsymbol{A}$ 的奇异值，也就是说，增加了 $n-r$ 个或 $m-r$ 个零奇异值. 在下面的定理中，我们就将这样做.

设 $\boldsymbol{A}\in \boldsymbol{C}^{m\times n}$. 现对 $\boldsymbol{A}$ 有一个微小扰动，因而我们实际观测到的是矩阵 $\boldsymbol{B}=\boldsymbol{A}+\boldsymbol{E}$. 下面的定理建立了 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的奇异值之

间的关系.

定理 2.10.3 设 A、B 和 E 均为 $m\times n$ 阵，$B=A+E$. 将 A 和 B 的奇异值分别记为

$$\sigma_1 \geqslant \cdots \geqslant \sigma_n,$$

$$\mu_1 \geqslant \cdots \geqslant \mu_n.$$

则

$$|\mu_i - \sigma_i| \leqslant \|E\|_2, i = 1,2,\cdots,n.$$

证明 注意到 $\sigma_1^2\geqslant\cdots\geqslant\sigma_n^2$ 为 A^*A 的特征值，而 $\mu_1^2\geqslant\cdots\geqslant\mu_n^2$ 为 B^*B 的特征值. 若用 $\lambda_1(C)\geqslant\cdots\geqslant\lambda_n(C)$ 表示 n 阶方阵 C 的特征值，用 $\varphi_1,\cdots,\varphi_n$ 表示 B^*B 对应于 $\mu_1^2\geqslant\cdots\geqslant\mu_n^2$ 的标准正交化特征向量，依定理 2.8.2 我们有

$$\mu_{i+1}^2 = \lambda_{i+1}(B^*B) = \max_{\substack{x\neq 0 \\ \varphi_j^* x=0 \\ j=1,\cdots,i}} \frac{x^*(B^*B)x}{x^*x}$$

$$\leqslant \max_{\substack{x\neq 0 \\ \varphi_j^* x=0 \\ j=1,\cdots,i}} \frac{x^*(A+E)^*(A+E)x}{x^*x}$$

$$= \max_{\substack{x\neq 0 \\ \varphi_j^* x=0 \\ j=1,\cdots,i}} \left(\frac{x^*A^*Ax}{x^*x} + \frac{x^*E^*Ex}{x^*x} + 2\frac{x^*E^*Ax}{x^*x} \right).$$

对上式第三项的分子用 Cauchy-Schwarz 不等式，得

$$|x^*E^*Ax|^2 \leqslant (x^*E^*Ex)(x^*A^*Ax),$$

于是

$$\mu_{i+1}^2 \leqslant \max_{\substack{x\neq 0 \\ \varphi_j^* x=0 \\ j=1,\cdots,i}} \left[\left(\frac{x^*A^*Ax}{x^*x} \right)^{1/2} + \left(\frac{x^*E^*Ex}{x^*x} \right)^{1/2} \right]^2$$

$$\leqslant \left[\max_{\substack{x\neq 0\\ \varphi_j^* x=0\\ j=1,\cdots,i}} \left(\frac{x^* A^* A x}{x^* x} \right)^{1/2} + \max_{\substack{x\neq 0\\ \varphi_j^* x=0\\ j=1,\cdots,i}} \left(\frac{x^* E^* E x}{x^* x} \right)^{1/2} \right]^2$$

$$= \left(\sigma_{i+1} + \lambda_{i+1}^{\frac{1}{2}}(E^* E) \right)^2$$

$$\leqslant (\sigma_{i+1} + \|E\|_2)^2, \qquad i = 1,2,\cdots,n-1.$$

对上面的讨论作明显修改，知上式对 $i=0$ 也成立. 于是我们证明了

$$\mu_i \leqslant \sigma_i - \|E\|_2, \qquad i = 1,2,\cdots,n. \quad (2.10.4)$$

另一方面，对 $A=B+(-E)$，重复上面的讨论，可得

$$\sigma_i \leqslant \mu_i + \|E\|_2, \qquad i = 1,\cdots,n. \quad (2.10.5)$$

综合(2.10.4)与(2.10.5)，就有

$$\sigma_i - \|E\|_2 \leqslant \mu_i \leqslant \sigma_i + \|E\|_2, \qquad i = 1,\cdots,n.$$

定理证毕.

第三章　{1}-逆

从第一章给出的各种广义逆的定义，我们不难看出，{1}-逆是最基本、最重要的一种广义逆，它是研究其它广义逆的基础. 本章将讨论{1}-逆的结构、性质以及在矩阵方程求解和投影阵表示中的应用. 至于其它方面的应用，将放在后续章节的有关地方.

§3.1　{1}-逆的结构

对任意 $m\times n$ 矩阵 $\boldsymbol{A}$，我们用 $\boldsymbol{A}^-$ 或 $\boldsymbol{A}^{(1)}$ 表示它的任意一个{1}-逆，而用 $\boldsymbol{A}\{1\}$表示 $\boldsymbol{A}$ 的{1}-逆的全体. 本节用两种方法来刻画{1}-逆的结构.

定理 3.1.1　设 $\boldsymbol{A}\in \boldsymbol{C}^{m\times n}$，其秩为 r，其相抵标准形为

$$\boldsymbol{A}=\boldsymbol{P}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{Q}, \tag{3.1.1}$$

这里 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 分别为 $m\times m$ 和 $n\times n$ 可逆阵. 则

$$\boldsymbol{A}^-=\boldsymbol{Q}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{pmatrix}\boldsymbol{P}^{-1}, \tag{3.1.2}$$

这里 $\boldsymbol{B}_{12}$、$\boldsymbol{B}_{21}$和 $\boldsymbol{B}_{22}$分别为 $r\times(m-r)$、$(n-r)\times r$ 和 $(n-r)\times(m-r)$的任意矩阵.

证明　设 $\boldsymbol{X}$ 为任意一个 $\boldsymbol{A}^-$，则有 $\boldsymbol{AXA}=\boldsymbol{A}$. 将(3.1.1)代入，我们得到

$$\boldsymbol{P}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{QXP}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{Q}=\boldsymbol{P}\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{Q}.$$

因为 P 和 Q 皆为可逆阵，上式等价于

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} QXP \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \qquad (3.1.3)$$

若记

$$QXP = B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \qquad (3.1.4)$$

则从(3.1.3)可推得 $B_{11}=I_r$. 这就证明了，X 为 A 的一个广义逆 A^-，当且仅当

$$QXP = \begin{pmatrix} I_r & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \qquad (3.1.5)$$

从此式立得(3.1.2). 定理证毕.

这个定理给出了 A^- 的一种刻画. 在(3.1.2)中，三个矩阵 B_{12}、B_{21} 和 B_{22} 可以任意选取. 因此，当 $r=r(A)<n$ 或 m 时，存在无穷多个 A^-. 仅当 $r(A)=m=n$ 时 A^- 才是唯一，且就是通常的 A^{-1}，将 B_{12}、B_{21} 和 B_{22} 任意变化，取遍所有可能取的值，我们就获得了全体的 A^-，即 $A\{1\}$. 于是

$$A\{1\} = \left\{A^-: A^- = Q^{-1}\begin{pmatrix} I_r & B_{12} \\ B_{21} & B_{22} \end{pmatrix}P^{-1},\right.$$

$$\left. B_{ij} \text{ 为适当阶数的任意阵}\right\}.$$

与前一定理不同，下面我们用一个特定的 A^- 来刻画 $A\{1\}$.

定理 3.1.2 设 $A \in C^{m\times n}$，A^- 为其一个特定的 $\{1\}$-逆，则

$$A\{1\} = \{A^- + U - A^-AUAA^-,$$
$$\text{其中 } U \in C^{n\times m} \text{ 任意阵}\}; \qquad (3.1.6)$$

$$A\{1\} = \{A^- + Z(I_m - AA^-) + (I_n - A^-A)Y,$$
$$\text{其中 } Z, Y \in C^{n\times m} \text{ 任意阵}\}. \qquad (3.1.7)$$

证明 将(3.1.6)和(3.1.7)右端的集合分别记为 S_1 和 S_2. 容易验证,这两个集合中的每个矩阵都满足 $\boldsymbol{AXA}=\boldsymbol{A}$,即 $S_i\subset\boldsymbol{A}\{1\}, i=1,2$.

反过来,对 $\boldsymbol{AXA}=\boldsymbol{A}$ 的任一解 $\boldsymbol{X}$,取 $\boldsymbol{U}=\boldsymbol{X}-\boldsymbol{A}^-$,便可推出 $\boldsymbol{X}\in S_1$,于是 $\boldsymbol{A}\{1\}\subset S_1$. 若取

$$\boldsymbol{Z}=\boldsymbol{X}-\boldsymbol{A}^-,$$
$$\boldsymbol{Y}=\boldsymbol{XAA}^-,$$

则有 $\boldsymbol{X}\in S_2$,于是 $\boldsymbol{A}\{1\}\subset S_2$. 定理证毕.

§3.2 基 本 性 质

设 λ 为一个数,定义

$$\lambda^+=\begin{cases}\lambda^{-1}, & 若\ \lambda\neq 0,\\ 0, & 若\ \lambda=0.\end{cases}$$

应用(3.1.2),不难证明{1}-逆的下列性质:

定理 3.2.1 设 $\boldsymbol{A}\in\boldsymbol{C}^{m\times n}$.

(a) $\boldsymbol{A}^-$ 总是存在的,$\boldsymbol{A}^-$ 唯一当且仅当 $\boldsymbol{A}$ 为可逆方阵,且此时 $\boldsymbol{A}^-=\boldsymbol{A}^{-1}$.

(b) $(\boldsymbol{A}^-)^*\in\boldsymbol{A}^*\{1\}$.

(c) $\lambda^+\boldsymbol{A}^-\in(\lambda\boldsymbol{A})\{1\}$.

(d) 若 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 分别为 $m\times m$、$n\times n$ 可逆阵,则

$$\boldsymbol{Q}^{-1}\boldsymbol{A}^-\boldsymbol{P}^{-1}\in(\boldsymbol{PAQ})\{1\}, \qquad (3.2.1)$$

对任一 $\boldsymbol{A}^-$ 成立.

这些事实的证明请读者去完成,但须注意结论(b)、(c)和(d)与 $\lambda\neq 0$,$\boldsymbol{A}$ 可逆时相对应结论的差别. 这些差别渊源于{1}-逆的不唯一性.

我们在§2.6已经定义了幂等阵. 即一个方阵 $\boldsymbol{A}$ 若满足

$A^2=A$,则称 A 为幂等阵,下面的定理给出了幂等阵和广义逆矩阵之间的一些关系.

定理 3.2.2 设 $A\in C^{m\times n}$,则

(a) $r(A^-)\geqslant r(A)$,对任一 A^- 成立.

(b) 对任意 A^-,A^-A 为幂等阵,且 $r(A^-A)=r(A)$. 进一步

$$A^-A=I_n\Longleftrightarrow r(A)=n. \tag{3.2.2}$$

(c) 对任意 A^-,AA^- 为幂等阵,且 $r(AA^-)=r(A)$. 进一步

$$AA^-=I_m\Longleftrightarrow r(A)=m. \tag{3.2.3}$$

证明 结论(a)容易从(3.1.2)推出.

(b) 利用{1}-逆的定义,容易验证 A^-A 的幂等性,再利用(3.1.2)可得

$$A^-A=P\begin{pmatrix} I_r & 0 \\ B_{21} & 0 \end{pmatrix}P^{-1},$$

于是 $r(A^-A)=r=r(A)$.

(3.2.2)的必要性部分是显然的. 下证充分性. 若 $r(A)=n$,则由已证部分知 $r(A^-A)=n$,这表明 A^-A 为可逆幂等阵. 用 $(A^-A)^{-1}$ 左乘 $(A^-A)^2=A^-A$,便得到 $A^-A=I_n$.

用类似的方法可证明结论(c). 定理证毕.

注 定理 3.2.2 的三条结论进一步显示了矩阵的广义逆与普通逆不同之处. 若 A 可逆,则 $A^{-1}A=AA^{-1}=I$. 注意 I 也是幂等阵,且是具有最大秩的幂等阵. 而当 A 不可逆时,A^-A 和 AA^- 都是幂等阵,它们的秩等于 A 的秩.

推论 3.2.1 对任意的矩阵 A,

$$\mathscr{M}(AA^-)=\mathscr{M}(A), \tag{3.2.4}$$

$$\mathscr{N}(A^-A)=\mathscr{N}(A). \tag{3.2.5}$$

这两个事实的证明是容易的. 因为显然有 $\mathscr{M}(AA^-)\subset\mathscr{M}(A)$，$\mathscr{N}(A^-A)\subset\mathscr{N}(A)$. 再结合定理 3.2.2 中的 $r(AA^-)=r(A)=r(AA^-)$，便得到(3.2.4)和(3.2.5).

对任意 $m\times n$ 矩阵 A，若 $r(A)=n$，则不难证明存在 $n\times m$ 阵 X，使得

$$XA = I_n \qquad (3.2.6)$$

我们称 X 为 A 的左逆阵. 类似地，若 $r(A)=m$，则可以证明，存在 $n\times m$ 阵 X，使得

$$AX = I_m \qquad (3.2.7)$$

我们称 X 为 A 的右逆阵. (3.2.2)表明，若 A 是列满秩(即 A 的秩等于 A 的列数)，则任一个 A^- 都是 A 的左逆阵. 类似地，(3.2.3)表明，若 A 是行满秩(即 A 的秩等于 A 的行数)，则任一个 A^- 都是 A 的右逆阵.

一般说来，A^- 是不唯一的，因而矩阵 BA^-C 与广义逆 A^- 的选择有关，也就是说，对不同的 A^-，矩阵 BA^-C 可以是不相同的. 但是，当矩阵 B、C 与 A 满足一定条件时，BA^-C 就可以与 A^- 的选择无关. 这个性质很重要，它为广义逆矩阵的许多应用奠定了基础.

定理 3.2.3 设 $B\neq 0$，$C\neq 0$. 则矩阵 BA^-C 与 A^- 选择无关，当且仅当

$$\mathscr{M}(C)\subset\mathscr{M}(A),\ \mathscr{M}(B')\subset\mathscr{M}(A'), \qquad (3.2.8)$$

其中 $\mathscr{M}(A)$ 表示 A 的列向量张成的子空间.

证明 充分性 若(3.2.8)成立，则存在矩阵 X 和 Y，使得 $C=AX$，$B'=A'Y$. 于是

$$BA^-C = Y'AA^-AX = Y'AX,$$

右端与 A^- 无关.

必要性 应用(3.1.6)，任意一个 A^- 可表为

$$A^- = A^{(1)} + U - A^{(1)}AUAA^{(1)},$$

其中 $A^{(1)}$表示任一固定的{1}-逆，于是

$$BA^- C = BA^{(1)}C + BUC - BA^{(1)}AUAA^{(1)}C.$$

若 BA^-C 与 A^-选择无关，必有 $BA^-C=BA^{(1)}C$，因而

$$BUC - BA^{(1)}AUAA^{(1)}C = 0. \qquad (3.2.9)$$

取 $U=A^{(1)}AZ$，其中 Z 为任意阵，则

$$BA^{(1)}AZ(C - AA^{(1)}C) = 0,$$

由 Z 的任意性可知，或

$$BA^{(1)}A = 0, \qquad (3.2.10)$$

或

$$C = AA^{(1)}C. \qquad (3.2.11)$$

若(3.2.10)成立，从(3.2.9)可推出 $BUC=0$，由 U 的任意性，我们有 $B=0$ 或 $C=0$，这与原假设矛盾. 于是(3.2.11)成立. 这就证明了 $\mathscr{M}(C)\subset\mathscr{M}(A)$.

若取 $U=ZAA^{(1)}$，用类似方法可证明 $\mathscr{M}(B')\subset\mathscr{M}(A')$. 定理证毕.

定理 3.2.4 $AA^-B=B$ 当且仅当 $\mathscr{M}(B)\subset\mathscr{M}(A)$.

证明 必要性是显然的，下证充分性. 若 $\mathscr{M}(B)\subset\mathscr{M}(A)$，则存在矩阵 X，使得 $B=AX$. 于是

$$AA^- B = AA^- AX = AX = B.$$

定理证毕.

定理 3.2.5 对任意的矩阵 A，

(a) $A(A^*A)^-A^*$ 与广义逆$(A^*A)^-$的选择无关，且为 Hermite 阵，其秩与 A 相同.

(b) $A(A^*A)^-A^*A=A$, $A^*A(A^*A)^-A^*=A^*$. (3.2.12)

(c) 设 V 为任一满足条件

$$r(A^*VA) = r(A) \qquad (3.2.13)$$

的 Hermite 阵，则

$$A(A^*VA)^- A^*VA = A, \qquad (3.2.14)$$

$$A^*VA(A^*VA)^- A^* = A^*. \qquad (3.2.15)$$

证明 (a) 因为 $\mathscr{M}(A^*)=\mathscr{M}(A^*A)$，故存在矩阵 X，使得 $A^*=A^*AX$. 因而

$$A(A^*A)^- A^* = X^*A^*A(A^*A)^- A^*AX = X^*A^*AX,$$

上式右端与 $(A^*A)^-$ 无关，且为 Hermite 阵. 关于秩的结论是显然的，因为在上式中我们可以取一个可逆的 $(A^*A)^-$.

(b) 是 (c) 中 $V=I$ 的特例，于是我们只证 (c). 由条件 (3.2.13) 知，对任意向量 x

$$A^*VAx = 0 \Rightarrow Ax = 0.$$

利用此事实，从

$$A^*V[A(A^*VA)^- A^*VA - A] = 0,$$

可得

$$A(A^*VA)^- A^*VA - A = 0.$$

(3.2.14) 得证，用类似方法可证明 (3.2.15). 定理证毕.

众所周知，若 A^{-1} 和 $(A+uv^*)^{-1}$ 皆存在，则

$$(A+uv^*)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1+v^*A^{-1}u},$$

其中 u 和 v 为向量. 与此相对应，我们有如下结果.

定理3.2.6 设 $A\in C^{n\times n}$，u 和 v 皆为 $n\times 1$ 向量，且 $u\in\mathscr{M}(A)$，$v\in\mathscr{M}(A^*)$. 则

$$A^- - \frac{A^-u\cdot v^*A^-}{1+v^*A^-u} \in (A+uv^*)\{1\}. \qquad (3.2.16)$$

证明 因为 $u\in\mathscr{M}(A)$，$v\in\mathscr{M}(A^*)$，从定理3.2.4，得

$$AA^-u = u,$$

$$v^*A^-A = v^*.$$

利用此事实不难验证所要的结论.

§3.3 矩阵方程的解

{1}-逆的重要应用之一是关于矩阵方程和线性方程组解的表示，它起着与通常逆矩阵完全类似的作用. 另一方面，本节的结果在各种广义逆的表征中起着重要作用(见第五章).

定理3.3.1 设 $A\in C^{m\times n}, B\in C^{p\times q}, H\in C^{m\times q}$. 则矩阵方程

$$AXB = H \tag{3.3.1}$$

相容，当且仅当存在 A^- 和 B^-，使得

$$AA^-HB^-B = H. \tag{3.3.2}$$

当(3.3.1)相容时，其通解为

$$X = A^-HB^- + Y - A^-AYBB^-, \tag{3.3.3}$$

其中 Y 为任意 $n\times p$ 矩阵，A^- 和 B^- 为任意广义逆.

证明 若存在 A^- 和 B^- 使(3.3.2)成立，显然 $X=A^-HB^-$ 为(3.3.1)的一个解. 反过来，若 X 为(3.3.1)的一个解，则对任意的 A^- 和 B^-，有

$$H = AXB = AA^-AXBB^-B = AA^-HB^-B,$$

(3.3.2)得证.

假设对某个 A^- 和 B^-，(3.3.2)成立，则容易验证(3.3.3)给出的 X 是(3.3.1)的解. 反过来，若 X 是(3.3.1)的解，则对任意的 A^- 和 B^-，X 可表为

$$X = A^-HB^- + X - A^-AXBB^-,$$

它具有(3.3.3)的形式. 定理证毕.

注 从定理证明过程知，若(3.3.1)相容，则对一切 A^- 和 B^-，(3.3.2)成立.

推论3.3.1 设 $A\in C^{m\times n}$，矩阵方程 $A^*XA=0$的通解为

$$X = Y - (AA^-)^*YAA^-,$$

这里 $Y\in C^{m\times m}$任意阵. 当 Y 为 Hermite 阵时，上式表示 Hermite 通解.

下面的推论是 X 和 H 为向量的特殊情形.

推论3.3.2 设 $A\in C^{m\times n}$，$b\in C^m$. 则线性方程组 $Ax=b$ 相容，当且仅当存在一个 A^-，使得

$$AA^-b = b. \tag{3.3.4}$$

若(3.3.4)成立，则 $Ax=b$ 的通解为

$$x = A^-b + (I - A^-A)t,$$

这里 $t\in C^n$ 的任意向量.

推论3.3.3 (a)设 $A\in C^{m\times n}$，$H\in C^{m\times q}$. 则矩阵方程

$$AX = H \tag{3.3.5}$$

相容，当且仅当存在 A^-使得

$$AA^-H = H. \tag{3.3.6}$$

当(3.3.5)相容时，通解为

$$X = A^-H + (I_n - A^-A)U, \tag{3.3.7}$$

其中 U 为 $n\times q$ 任意阵，A^-为任一固定的{1}-逆.

(b) 设 $A\in C^{m\times n}$，$H\in C^{q\times n}$. 则矩阵方程

$$XA = H \tag{3.3.8}$$

相容，当且仅当存在 A^-使得

$$HA^-A = H. \tag{3.3.9}$$

当矩阵方程(3.3.8)相容时，其通解可表为

$$X = HA^- + W(I_m - AA^-),$$

这里 W 为 $q\times m$ 任意阵.

推论3.3.3的两条结论分别是定理3.3.1中 $A=I$ 和 $B=I$ 的两种特殊情形.

推论3.3.4 设 $A\in C^{m\times n}$，则矩阵方程 $AX=0$的通解为

$$X=(I_n-A^- A)U,$$

其中 U 为任意阵. 而矩阵方程 $XA=0$的通解为

$$X=W(I_m-AA^-),$$

其中 W 为任意阵.

定理3.3.2 设 $A\in C^{m\times n}$，$G\in C^{n\times m}$，则 $G\in A\{1\}$当且仅当对一切相容线性方程组 $Ax=b$，$x=Gb$ 是一个解.

证明 必要性 设 $Ax=b$ 相容，由推论3.3.2知，存在一个 A^- 使得(3.3.4)成立. 于是对任意 $G\in A\{1\}$，

$$AGb=AGAA^- b=AA^- b=b,$$

必要性得证.

充分性 记 $A=(a_1,\cdots,a_n)$. 显然

$$Ax=a_j,\quad j=1,\cdots,n$$

都相容，若 Ga_j 是 $Ax=a_j$ 的解，

$$AGa_j=a_j,\quad j=1,\cdots,n.$$

等价地，$AGA=A$. 即 $G\in A\{1\}$. 定理证毕.

定理3.3.3 设 A、B、C 和 D 分别为 $m\times n$、$m\times p$、$p\times q$ 和 $n\times q$ 矩阵. 则矩阵方程 $AX=B$ 和 $XC=D$ 有公共解当且仅当每个方程都有解，且 $AD=BC$.

证明 假设 $AX=B$ 与 $XC=D$ 都有解，则对任意的 A^- 和 C^-，有

$$AA^- B=B,\quad DC^- C=D.$$

若 $AD=BC$，则易验证

$$X=A^- B+DC^- -A^- ADC^- \tag{3.3.10}$$

是 $AX=B$ 和 $XC=D$ 的公共解.

反过来，设 X_0为它们的公共解，则

$$AD=AX_0C=BC.$$

定理证毕.

定理3.3.4 设 A、B、C 和 D 分别为 $m\times n$、$m\times p$、$p\times q$ 和 $n\times q$ 矩阵，且 $AX=B$ 和 $XC=D$ 有公共解 X_0. 则

$$X = X_0 + (I - A^- A)Y(I - CC^-) \qquad (3.3.11)$$

为公共通解，这里 A^- 和 C^- 为 A 和 C 的任意一个{1}-逆，Y 为适当阶数的任意阵.

证明 因为 $AX_0=B$，$X_0C=D$，故对(3.3.11)给出的每个 X 都有

$$AX = AX_0 = B,$$
$$XC = X_0C = D,$$

即(3.3.11)给出的每个 X 都是 $AX=B$ 和 $XC=D$ 的公共解.

反过来，对 $AX=B$ 和 $XC=D$ 的每个公共解 X，在(3.3.11)中取 $Y=X-X_0$，定理证毕.

注 从(3.3.10)知，在(3.3.11)中可取

$$X_0 = A^- B + DC^- - A^- ADC^-.$$

于是，若 $AX=B$ 和 $XC=D$ 都有解且满足 $AD=BC$. 则

$$X = A^- B + DC^- - A^- ADC^- + (I - A^- A)Y(I - CC^-) \qquad (3.3.12)$$

为公共通解.

推论3.3.5 设 A 和 B 分别为 $m\times n$ 和 $p\times q$ 矩阵，则矩阵方程 $AX=0$ 和 $XB=0$ 的公共通解为

$$X = (I_n - A^- A)Y(I_p - BB^-),$$

其中 Y 为 $n\times p$ 的任意阵.

§3.4 投影阵的表示定理

在§2.6，我们讨论了投影阵和正交投影阵的基本性质，

证明了一个方阵为投影阵的充要条件是幂等阵，而为正交投影阵的充要条件是为 Hermite 幂等阵. 但是，在那里，我们没有构造性地给出投影阵和正交投影阵. 本节我们利用{1}-逆建立投影阵的表示.

设 $M \in C^{n\times n}$，且 $M>0$，即 M 为正定 Hermite 阵. 在 C^n 中定义内积

$$(x,y)_M = y^* Mx. \tag{3.4.1}$$

设 $A\in C^{n\times p}$. 记 $A_M^{\perp}$ 为满足条件 $A^* MA_M^{\perp}=0$ 且具有最大秩的矩阵. 设 $B\in C^{n\times q}$，假定

$$C^n = \mathscr{M}(A) \oplus \mathscr{M}(B). \tag{3.4.2}$$

为符号简单计，我们把投影阵 $P_{\mathscr{M}(A),\mathscr{M}(B)}$ 简记为 $P_{A,B}$，而把正交投影阵 $P_{\mathscr{M}(A)}$ 记为 P_A.

定理3.4.1 设 $A\in C^{n\times p}$，$B\in C^{n\times q}$，且关系式(3.4.2)成立. 则

$$P_{A,B} = A(ZMA)^- ZM, \tag{3.4.3}$$

其中 $Z=(B_M^{\perp})^*$.

证明 为简单计，在本定理证明中，将 $P_{A,B}$ 记为 P. 根据定义，P 是下列矩阵方程的解：

$$PA = A, \tag{3.4.4}$$

$$PB = 0. \tag{3.4.5}$$

将(3.4.5)改写为

$$(\overline{B})^* M(M^{-1}P') = 0,$$

因而

$$\mathscr{M}(M^{-1}P') \subset \mathscr{M}((\overline{B})_M^{\perp}).$$

故存在矩阵 K，使得 $M^{-1}P' = (\overline{B})_M^{\perp}K$. 从此矩阵方程可解得

$$P = K\{(\overline{B})_M^{\perp}\}' M. \tag{3.4.6}$$

代入(3.4.4),得

$$K\left((\overline{B})_M^{\perp}\right)'MA = A.$$

从此式再解出 K,我们有

$$K = A((\overline{B}_M)^{\perp})'MA)^{-}.$$

代入(3.4.6)

$$\begin{aligned} P &= A((((\overline{B}_M)^{\perp})'MA)^{-}\ ((\overline{B})_M^{\perp})'M \\ &= A(ZMA)^{-}\ ZM. \end{aligned}$$

定理证毕.

对于正交投影阵 P_A,因为 $\mathscr{M}(B)=\mathscr{M}(A)^{\perp}=\mathscr{M}(A_M^{\perp})$ 此时$(B_M^{\perp})^{*}=A^{*}$,于是从定理3.4.1我们立即得到

推论3.4.1 设 $A\in C^{n\times p}$.

(a) 若内积定义为(3.4.1),则向 $\mathscr{M}(A)$的正交投影阵为

$$P_A = A(A^{*}MA)^{-}A^{*}M, \tag{3.4.7}$$

且 P_A 与所含广义逆的选择无关.

(b) 若在(3.4.1)中,$M=I$. 则向 $\mathscr{M}(A)$的正交投影阵

$$P_A = A(A^{*}A)^{-}A^{*}. \tag{3.4.8}$$

且 P_A 与所含广义逆的选择无关.

§3.5 具有给定秩的{1}-逆

设 $A\in C^{m\times n}$. 我们知道 $r(A^{-})\geqslant r(A)$,且 $r(A^{-})$可以取任意正整数 k,这里 k 满足

$$r(A)\leqslant k\leqslant \min\{m,n\}. \tag{3.5.1}$$

一个有意义的问题是:对给定的满足(3.5.1)的 k,如何刻画具有给定秩 k 的 A^{-}. 本节就来回答这个问题.

引理3.5.1 设 H 为 $n\times n$ 幂等阵. 则存在正定 Hermite

阵 $\boldsymbol{P}$,使得 $\boldsymbol{P}^{-1}\boldsymbol{H}^*\boldsymbol{P}=\boldsymbol{H}$.

证明 设 $r(\boldsymbol{H})=r$,则 $\boldsymbol{I}-\boldsymbol{H}$ 为幂等阵且秩为 $n-r$. 考虑 $\boldsymbol{H}$ 和 $\boldsymbol{I}-\boldsymbol{H}$ 的满秩分解(关于满秩分解,见定理2.4.5)

$$\boldsymbol{H} = \boldsymbol{A}_1\boldsymbol{B}_1,$$
$$\boldsymbol{I} - \boldsymbol{H} = \boldsymbol{A}_2\boldsymbol{B}_2,$$

这里 $\boldsymbol{A}_1 \in \boldsymbol{C}^{n\times r}$, $\boldsymbol{B}_1 \in \boldsymbol{C}^{r\times n}$, 且 $r(\boldsymbol{A}_1) = r(\boldsymbol{B}_1) = r$, 而 $\boldsymbol{A}_2 \in \boldsymbol{C}^{n\times(n-r)}$, $\boldsymbol{B}_2 \in \boldsymbol{C}^{(n-r)\times n}$, 且 $r(\boldsymbol{A}_2)=r(\boldsymbol{B}_2)=n-r$. 记

$$\boldsymbol{A} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2),$$

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{pmatrix}.$$

注意,$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 皆为 $n\times n$ 方阵,且

$$\boldsymbol{AB} = \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_2 = \boldsymbol{H} + (\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{I}_n,$$

于是 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 皆为可逆阵,且

$$\boldsymbol{A} = \boldsymbol{B}^{-1} = (\boldsymbol{B}^*\boldsymbol{B})^{-1}\boldsymbol{B}^* \stackrel{d}{=} \boldsymbol{P}^{-1}\boldsymbol{B}^*,$$

这里 $\boldsymbol{P}=\boldsymbol{B}^*\boldsymbol{B}>0$. 从上式可得

$$\boldsymbol{A} = (\boldsymbol{A}_1 \vdots \boldsymbol{A}_2) = (\boldsymbol{P}^{-1}\boldsymbol{B}_1^* \vdots \boldsymbol{P}^{-1}\boldsymbol{B}_2^*),$$

故

$$\boldsymbol{A}_1 = \boldsymbol{P}^{-1}\boldsymbol{B}_1^*,$$
$$\boldsymbol{B}_1 = \boldsymbol{A}_1^*\boldsymbol{P}^* = \boldsymbol{A}_1^*\boldsymbol{P}.$$

最后我们得到

$$\boldsymbol{H} = \boldsymbol{A}_1\boldsymbol{B}_1 = \boldsymbol{P}^{-1}\boldsymbol{B}_1^*\boldsymbol{A}_1^*\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{H}^*\boldsymbol{P}.$$

定理3.5.1 设 $\boldsymbol{A} \in \boldsymbol{C}^{m\times n}$ 且秩为 r, 正整数 k 满足(3.5.1). 则 $n\times m$ 矩阵 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的一个秩为 k 的{1}-逆,当且仅当 $\boldsymbol{G}$ 为 $\boldsymbol{A}+\boldsymbol{MN}$ 的一个{1,2}-逆,其中 $\boldsymbol{M}\in \boldsymbol{C}^{m\times(k-r)}$, $\boldsymbol{N}\in \boldsymbol{C}^{(k-r)\times n}$, 且

$$r(\boldsymbol{A} \vdots \boldsymbol{M}) = r\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{N} \end{pmatrix} = k. \tag{3.5.2}$$

证明 必要性 设 G 为 A 的一个{1}-逆，$r(G)=k$，则 $H=GA$ 是幂等阵，且 $r(H)=r$，根据引理3.5.1知，存在正定阵 M，使得

$$M^{-1}H^{*}M = H.$$

于是

$$MH = H^{*}M = (MH)^{*},$$

即 MH 为 Hermite 阵. 假定在 C^{n} 中定义内积 $(x,y)_{M}=y^{*}Mx$，根据定理2.6.7，由 H 是幂等阵，MH 为 Hermite 阵，可推出 $H=GA$ 为向 $\mathscr{M}(GA)$ 上的正交投影阵.

记

$$P_{G} = G(G^{*}MG)^{-}G^{*}M,$$

这是在同样内积下向 $\mathscr{M}(G)$ 上的正交投影阵. 若定义

$$U = (G^{*}MG)^{-}G^{*}M - A,$$

则 $GU=P_{G}-GA$ 是向 $\mathscr{M}(GA)$ 在 $\mathscr{M}(G)$ 内的正交补空间的正交投影阵，因而 GU 也是幂等阵，

$$r(GU) = r(P_{G}) - r(GA) = k - r.$$

记 $X=UGU$，则 $GX=GU$，这表明 GX 也是一个幂等阵，

$$r(X) = r(GU) = k - r.$$

再注意到

$$G(A+X) = P_{G}, \tag{3.5.3}$$

于是

$$\begin{aligned} r(G(A+X)) &= r(G) = k = r + (k-r) \\ &= r(A) + r(X) = r(A+X). \end{aligned} \tag{3.5.4}$$

(3.5.3)与(3.5.4)表明，$G(A+X)$ 为幂等阵且 $G(A+X)$ 与 $A+X$ 有相同的秩. 据此可推出 G 为一个 $(A+X)^{-}$. 再利用定理5.1.2(因为 $r(G)=r(A+X)$)知 G 为一个 $(A+X)^{(1,2)}$. 最

后对 X 作满秩分解 $X=MN$，其中 $M\in C^{m\times(k-r)}$，$N\in C^{(k-r)\times n}$，它们的秩皆为 $k-r$. 必要性得证.

充分性　设 $A=M_1N_1$ 是 A 的满秩分解，其中 $M_1\in C^{m\times r}$，$N_1\in C^{r\times n}$，且秩皆为 r，利用(3.5.2)得

$$A+MN=(M_1\ \vdots\ M)\begin{pmatrix}N_1\\N\end{pmatrix}$$

是 $A+MN$ 的满秩分解. 将 $(A+MN)^{(1,2)}$ 分解为

$$(A+MN)^{(1,2)}=(Q_1\ \vdots\ Q_2)\begin{pmatrix}P_1\\P_2\end{pmatrix}=Q_1P_1+Q_2P_2,$$

其中 $(Q_1\ \vdots\ Q_2)$ 为 $\begin{pmatrix}N_1\\N\end{pmatrix}$ 的右逆，$\begin{pmatrix}P_1\\P_2\end{pmatrix}$ 为 $(M_1\ \vdots\ M)$ 的左逆，$Q_1\in C^{n\times r}$，$Q_2\in C^{n\times(k-r)}$，$P_1\in C^{r\times m}$，$P_2\in C^{(k-r)\times m}$. 故

$$\begin{pmatrix}N_1\\N\end{pmatrix}(Q_1P_1+Q_2P_2)(M_1\quad M)=I_k.$$

从而有

$$N_1(Q_1P_1+Q_2P_2)M_1=I_r,$$
$$M_1N_1(Q_1P_1+Q_2P_2)M_1N_1=M_1N_1.$$

这后一式表明

$$A(A+MN)^{(1,2)}A=A,$$

即 $(A+MN)^{(1,2)}$ 是一个 A^-，它的秩等于 $A+MN$ 的秩 k. 定理证毕.

这个定理从理论上刻画了具有给定秩的{1}-逆的性质. 但如果我们想找出一个具体的{1}-逆，下面的方法是简单易行的.

设 $A\in C^{m\times n}$，且它有相抵标准形

$$A=P\begin{pmatrix}I_r&0\\0&0\end{pmatrix}Q,$$

这里 $P\in C^{m\times m}$, $Q\in C^{n\times n}$, 皆为可逆阵, $r=r(A)$. 对满足(3.5.1)的 k,取

$$G=Q^{-1}\begin{pmatrix}I_{rk} & 0\\ 0 & 0\end{pmatrix}P\ ,$$

容易验证,G 是一个 A 且 $r(G)=k$.

§ 3.6 具有给定列空间与零空间的{1}-逆

设 $A\in C^{m\times n}$. 记 $L=\mathscr{R}(A)$, $M=\mathscr{N}(A)$. 因为对任一 A , $A\quad A$ 和 AA 都是幂等阵,所以由定理2.6.2和推论2.6.1知,它们都是投影阵,且

$$A\quad A=P_{\mathscr{R}(A\quad A),\mathscr{N}(A\quad A)}, \tag{3.6.1}$$

$$AA\quad =P_{\mathscr{R}(AA\quad),\mathscr{N}(AA\quad)}. \tag{3.6.2}$$

不难证明

$$\mathscr{R}(AA\quad)=\mathscr{R}(A),$$

$$\mathscr{N}(A\quad A)=\mathscr{N}(A).$$

再记

$$T=\mathscr{R}(A\quad A), \tag{3.6.3}$$

$$S=\mathscr{N}(AA\quad). \tag{3.6.4}$$

则(3.6.1)和(3.6.2)可改写为

$$A\quad A=P_{T,M}, \tag{3.6.5}$$

$$AA\quad=P_{L,S}. \tag{3.6.6}$$

对任意给定的矩阵 A,子空间 L 和 M 是完全确定的. 我们的问题是,对满足

$$L\oplus S=C^{n}, \tag{3.6.7}$$

$$T\oplus M=C^{m}. \tag{3.6.8}$$

的两个子空间 S 和 T，是否存在 A 使得(3.6.5)和(3.6.6)同时成立?等价地，使得(3.6.3)和(3.6.4)同时成立?

为了回答这个问题，我们先证明如下引理：

引理3.6.1 (a) $P_{S_1,S_2}A=A \Longleftrightarrow \mathscr{M}(A)\subset S_1$.

(b) $AP_{S_1,S_2}=A \Longleftrightarrow S_2\subset \mathscr{N}(A)$.

证 注意到事实

$$\mathscr{M}(P_{S_1,S_2}) = S_1,$$

(a)的证明是容易的. 现在证明(b). 因为

$$\mathscr{M}(I - P_{S_1,S_2}) = \mathscr{N}(P_{S_1,S_2}) = S_2, \qquad (3.6.9)$$

若

$$AP_{S_1,S_2} = A,$$

必有

$$A(I - P_{S_1,S_2}) = 0,$$

于是

$$S_2 = \mathscr{M}(I - P_{S_1,S_2}) \subset \mathscr{N}(A).$$

反过来，若 $S_2\subset \mathscr{N}(A)$. 则由(3.6.9)得

$$0 = A(I - P_{S_1,S_2}) = A - AP_{S_1,S_2},$$

引理证毕.

定理3.6.1 设 $A\in C^{n\times m}$，$L=\mathscr{M}(A)$，$M=\mathscr{N}(A)$，S 和 T 分别为 C^n 和 C^m 的两个子空间，且满足(3.6.7)和(3.6.8). 则满足(3.6.3)和(3.6.4)的广义逆 A^- 具有如下形式：

$$A^- = P_{T,M}A^{(1)}P_{L,S} + (I_m - A^{(1)}A)Y(I_n - AA^{(1)}), \qquad (3.6.10)$$

其中 $A^{(1)}$ 为任意固定的一个广义逆，Y 为 $m\times n$ 任意阵.

证明 问题等价于证明(3.6.10)是矩阵方程(3.6.5)和(3.6.6)的公共解. 根据引理3.6.1，得

$$P_{L,S}A = A = AP_{T,M}.$$

于是从定理3.3.3知，矩阵方程(3.6.5)和(3.6.6)有公共解.再结合定理3.3.4，我们只需证明

$$X = P_{T,M}A^{(1)}P_{L,S}$$

是(3.6.5)和(3.6.6)的一个解.

事实上，再次利用引理3.6.1，得

$$XA = P_{T,M}A^{(1)}P_{L,S}A = P_{T,M}A^{(1)}A = P_{T,M},$$

$$AX = AP_{T,M}A^{(1)}P_{L,S} = AA^{(1)}P_{L,S} = P_{L,S}.$$

这就完成了所要的证明. 定理证毕.

从实用的角度看，(3.6.10)并不简单. 下一个定理往往更实用一些. 我们先证明一个预备事实.

引理3.6.2 设 $A \in C^{n \times p}$，$B \in C^{p \times q}$. 则

(a) $AB(AB)^- A = A \Longleftrightarrow r(AB) = r(A)$.

(b) $B(AB)^- AB = B \Longleftrightarrow r(AB) = r(B)$.

证明 (a) 因为 $AB(AB)^-$ 为向 $\mathscr{M}(AB)$ 的投影阵，依引理3.6.1(a)，得

$$AB(AB)^- A = A \Longleftrightarrow \mathscr{M}(A) \subset \mathscr{M}(AB).$$

因为 $\mathscr{M}(AB) \subset \mathscr{M}(A)$，故上式等价于

$$\mathscr{M}(A) = \mathscr{M}(AB)$$

$$\Longleftrightarrow r(A) = r(AB).$$

(b) 因为

$$(AB)^- AB = P_{\mathscr{M}((AB)^- AB), \mathscr{N}((AB)^- AB)}$$

$$= P_{\mathscr{M}((AB)^- AB), \mathscr{N}(AB)},$$

利用引理3.6.1(b)，得

$$B(AB)^- AB = B \Longleftrightarrow \mathscr{N}(AB) \subset \mathscr{N}(B).$$

注意到 $\mathscr{N}(B) \subset \mathscr{N}(AB)$，故上式等价于

$$\mathscr{N}(AB) = \mathscr{N}(B)$$

$$\Longleftrightarrow r(AB) = r(B).$$

引理证毕.

定理3.6.2 设 $A \in C^{m \times n}, P \in C^{n \times p}, Q \in C^{q \times m}$. 记

$$X = P(QAP)^- Q.$$

则 $X \in A\{1\} \Longleftrightarrow r(QAP) = r(A)$.

证明 充分性 因为

$$r(A) = r(QAP) \leqslant r(AP) \leqslant r(A),$$

于是 $r(AP) = r(A)$. 因而存在矩阵 $Y \in C^{p \times n}$, 使得 $A = APY$. 应用引理3.6.2, 得

$$AXA = AP(QAP)^- QA = AP(QAP)^- QAPY = APY = A.$$

必要性 若 $X \in A\{1\}$, 则

$$A = AXA = AXAXA = AP(QAP)^- QAP(QAP)^- QA,$$

因而

$$r(A) \leqslant r(QAP) \leqslant r(A),$$

必要性得证, 定理证毕.

注 易见, 当条件 $r(QAP) = r(A)$ 成立时, $X \in A\{1\}$, 而且满足 $\mathscr{M}(X) \subset \mathscr{M}(P)$, $\mathscr{N}(X) \supset \mathscr{N}(Q)$.

第四章 Moore-Penrose 广义逆

对任意矩阵 A,在第一章我们引进了 Moore-Penrose 广义逆,记之为 A^+,它是 Penrose 方程

$$
\begin{aligned}
&(1)\ AXA=A,\\
&(2)\ XAX=X,\\
&(3)\ (AX)^*=AX,\\
&(4)\ (XA)^*=XA
\end{aligned}
\tag{4.0.1}
$$

的解 X. 显然,A^+是一个 A^-,因此在第三章中我们所得到的关于 A^-的所有结论对 A^-也成立. 本章将系统论述 A^+的进一步性质.

§4.1 存在性及构造

本节我们证明 Moore-Penrose 广义逆的存在性和唯一性,并给出两种构造这一广义逆的方法.

定理4.1.1 对任意 $A\in C^{m\times n}$,若 Moore-Penrose 逆存在,则最多只有一个.

证明 设 $X,Y\in A\{1,2,3,4\}$,则反复利用 Penrose 四个方程,有

$$
\begin{aligned}
X &= XAX = X(AX)^* = XX^*A^* = XX^*A^*X^*A^*\\
&= XX^*A^*Y^*A^* = X(AX)^*(AY)^* = XAXAY = XAY\\
&= XAYAY = (XA)^*(YA)^*Y = A^*X^*A^*Y^*Y\\
&= A^*Y^*Y = (YA)^*Y = YAY = Y.
\end{aligned}
$$

定理证毕.

下面的两个定理给出了构造 A^+ 的两种方法，一种是基于 A 的奇异值分解，另一种是基于 A 的满秩分解，它们回答了 A^+ 的存在性问题.

定理4.1.2 设 $A\in C^{m\times n}$ 有奇异值分解

$$A=P\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}Q^*, \tag{4.1.1}$$

其中 $P\in C^{m\times m}$，$Q\in C^{n\times n}$，且皆为酉阵，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_r)$，$\sigma_i>0$，$i=1,\cdots,r$，$r=r(A)$. 则

$$A^+=Q\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{pmatrix}P^*. \tag{4.1.2}$$

证明 可以直接验证(4.1.2)右端满足 Penrose 方程(4.0.1)，定理证毕.

定理4.1.3 设 $A\in C^{m\times n}$ 有满秩分解

$$A=FG, \tag{4.1.3}$$

这里 $F\in C^{m\times r}$，$G\in C^{r\times n}$，$r=r(A)=r(F)=r(G)$. 则

$$A^+=G^*(F^*AG^*)^{-1}F^*. \tag{4.1.4}$$

证明 我们首先证明 F^*AG^* 为可逆阵. 从(4.1.3)得

$$F^*AG^*=F^*FGG^*,$$

注意到 $r(F^*F)=r(F)=r$，$r(GG^*)=r(G)=r$，于是方阵 F^*F 和 GG^* 皆为可逆阵，因而 F^*AG^* 也是可逆阵，且

$$(F^*AG^*)^{-1}=(GG^*)^{-1}(F^*F)^{-1}.$$

故有

$$G^*(F^*AG^*)^{-1}F^*=G^*(GG^*)^{-1}(F^*F)^{-1}F^*. \tag{4.1.5}$$

不难验证，上式右端满足 Penrose 方程(4.0.1). 定理证毕.

已经证明的三个定理表明，对任何矩阵 A，A^+ 存在且唯一. 我们把这一事实写成如下推论：

推论4.1.1 对任何矩阵 A，A^+ 存在且唯一.

推论4.1.2 在定理4.1.3假设下

$$A^+ = (FG)^+ = G^+ F^+. \tag{4.1.6}$$

证明 事实上我们可以直接验证：

$$G^+ = G^*(GG^*)^{-1},$$

$$F^+ = (F^*F)^{-1}F^*.$$

于是从(4.1.4)和(4.1.5)得到(4.1.6). 证毕.

需要指出，一般说来 $(AB)^+ \neq B^+A^+$，这是 Moore-Penrose 逆与普通逆矩阵的一个不同之处. 例如

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

此时

$$(AB)^+ = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}^+ = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

但是

$$B^+A^+ = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

下面的推论是定理4.1.3的特殊情况.

推论4.1.3 设 $a \neq 0, b \neq 0$，则

$$(ab^*)^+ = \frac{ba^*}{a^*ab^*b}.$$

推论4.1.4 设 $a \neq 0$，则

$$a^+ = \frac{a^*}{a^*a}.$$

最后一个定理是用{1}-逆来表征 Moore-Penrose 逆.

定理4.1.4 对任意矩阵 A，

$$A^+ = A^*(AA^*)^- A(A^*A)^- A^*.$$

利用定理3.2.5容易验证上式右端满足(4.0.1),请读者自己完成这个证明.

§4.2 基本性质

下面的两个定理给出了 $\boldsymbol{A}^+$ 的一些基本性质,其证明可以利用 $\boldsymbol{A}^+$ 的定义或定理4.1.2推出.

定理4.2.1 对任意的矩阵 $\boldsymbol{A}$,

(a) 若 $\boldsymbol{A}$ 可逆,则 $\boldsymbol{A}^{-1}=\boldsymbol{A}^+$.

(b) $(\boldsymbol{A}^+)^+=\boldsymbol{A}$.

(c) $(\boldsymbol{A}^*)^+=(\boldsymbol{A}^+)^*$, $(\boldsymbol{A}')^+=(\boldsymbol{A}^+)'$.

(d) 记

$$\lambda^+=\begin{cases}\lambda^{-1}, & 若\ \lambda\neq 0,\\ 0, & 若\ \lambda=0,\end{cases}$$

则

$$(\lambda\boldsymbol{A})^+=\lambda^+\boldsymbol{A}^+.$$

(e) 若 $\boldsymbol{D}=\mathrm{diag}(d_1,\cdots,d_n)$,则 $\boldsymbol{D}^+=\mathrm{diag}(d_1^+,\cdots,d_n^+)$.

注 定理4.2.1的(c)表明,若 $\boldsymbol{A}$ 为 Hermite 阵,则 $\boldsymbol{A}^-$ 也是 Hermite 阵.

定理4.2.2 对任意的矩阵 $\boldsymbol{A}$,

(a) $\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^*(\boldsymbol{A}^+)^*=(\boldsymbol{A}^+)^*\boldsymbol{A}^*\boldsymbol{A}$.

(b) $\boldsymbol{A}^+=\boldsymbol{A}^+(\boldsymbol{A}^+)^*\boldsymbol{A}^*=\boldsymbol{A}^*(\boldsymbol{A}^+)^*\boldsymbol{A}^+$.

(c) $(\boldsymbol{A}^*\boldsymbol{A})^+=\boldsymbol{A}^+(\boldsymbol{A}^+)^*$.

(d) $\boldsymbol{A}^+=(\boldsymbol{A}^*\boldsymbol{A})^+\boldsymbol{A}^*=\boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*)^+$.

推论4.2.1 假设 $\boldsymbol{A}=\sum_{i=1}^{k}\boldsymbol{A}_i$,且 $\boldsymbol{A}_i$ 满足

$$\boldsymbol{A}_i^*\boldsymbol{A}_j=\boldsymbol{0},\quad \boldsymbol{A}_i\boldsymbol{A}_j^*=\boldsymbol{0},$$

对一切 $i\neq j$，则

$$A^{+}=\sum_{i=1}^{k}A_i^{+}.$$

定理4.2.3 设 $A\in C^{m\times n}$，则

(a) A^+A, AA^+, I_n-A^+A 和 I_m-AA^+ 都是 Hermite 幂等阵.

(b) $r(A)=r(A^+)=r(A^-A)=r(AA^+)$,

$$r(I_n-A^+A)=n-r(A),$$
$$r(I_m-AA^+)=m-r(A).$$

(c) $I_n\geqslant A^+A, I_m\geqslant AA^+$，这里 $A\geqslant B$ 表示 $A-B\geqslant 0$.

利用定理4.1.2容易证明(a)和(b)的第一条，其余结论容易从(a)推出.

根据推论2.6.2，在标准内积下，每个 Hermite 幂等阵都是正交投影阵，于是 A^+A, AA^+, I_n-A^+A 和 I_m-AA^+ 都是正交投影阵.

为了进一步认识 AA^+ 和 A^+A 的本质并获得一些重要结果，我们将对这两个矩阵作进一步深入讨论.

若在 C^m 中定义了标准内积，由推论3.4.1知，向 $\mathscr{M}(A)$ 的正交投影阵为

$$P_A=A(A^*A)^-A^*. \tag{4.2.1}$$

因为 P_A 与所含的广义逆选择无关，并利用定理4.2.2(d)有

$$P_A=A(A^*A)^+A^*=AA^+. \tag{4.2.2}$$

于是矩阵 AA^+ 为向 $\mathscr{M}(A)$ 的正交投影阵. 结合(3.6.2)我们得到

$$AA^+=P_{\mathscr{M}(AA^+),\mathscr{N}(AA^+)}. \tag{4.2.3}$$

综上讨论，我们可得到如下定理：

定理4.2.4 对任意 $A\in C^{m\times n}$，

(a) $AA^+=P_A$.

(b) $\mathscr{M}(AA^+)\oplus\mathscr{N}(AA^+)=C^m$.

(c) $\mathscr{M}(AA^+)=\mathscr{M}(AA^*)=\mathscr{M}(A)$.

(d) $\mathscr{N}(AA^+)=\mathscr{N}(AA^*)=\mathscr{N}(A^+)=\mathscr{N}(A^*)$.

(e) $\mathscr{N}(AA^+)=\mathscr{M}(A)^\perp$.

与上面的结论完全相类似，向 $\mathscr{M}(A^*)$的正交投影阵为

$$P_{A^*} = A^*(AA^*)^- A = A^*(AA^*)^+ A = A^+ A, \tag{4.2.4}$$

因而

$$A^+ A = P_{\mathscr{M}(A^+A),\mathscr{N}(A^+A)}. \tag{4.2.5}$$

于是我们有与定理4.2.4完全平行的结论.

定理4.2.5 对任意 $A\in C^{m\times n}$，

(a) $A^+A=P_{A^*}$.

(b) $\mathscr{M}(A^+A)\oplus\mathscr{N}(A^+A)=C^n$.

(c) $\mathscr{M}(A^+A)=\mathscr{M}(A^*A)=\mathscr{M}(A^+)=\mathscr{M}(A^*)$.

(d) $\mathscr{N}(A^+A)=\mathscr{N}(A^*A)=\mathscr{N}(A)$.

(e) $\mathscr{N}(A)=\mathscr{M}(A^*)^\perp$.

众所周知，当 A、P、Q 皆可逆时，

$$(PAQ)^{-1} = Q^{-1}A^{-1}P^{-1}.$$

对{1}-逆，在前面我们也已证明

$$Q^{-1}A^- P^{-1} \in (PAQ)\{1\}.$$

但一般说来，对 Moore-Penrose 逆，$(PAQ)^+=Q^{-1}A^+P^{-1}$不必成立. 然而我们有下面的结论：

定理4.2.6 设 $A\in C^{m\times n}$.

(a) 若 P 和 Q 分别为 $m\times m$ 和 $n\times n$ 酉阵，则

$$(PAQ)^+= Q^{-1}A^+ P^{-1} = Q^* A^+ P^*.$$

(b) 若 P 和 Q 分别为 $k\times m$ 和 $l\times n$ 阵，且满足 $P^*P=I_m$，

$Q^*Q=I_n$. 则

$$(PAQ^*)^+=QA^+P^*.$$

本定理的证明是简单的，故略去.

我们知道，正规阵可以通过酉阵将其对角化. 因此利用定理4. 2. 6可以得到如下推论：

推论4. 2. 2 设 n 阶正规阵 A 有分解

$$A=U\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n\end{pmatrix}U^*,$$

其中 U 为 $n\times n$ 酉阵. 则

$$A^+=U\begin{pmatrix}\lambda_1^+ & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n^+\end{pmatrix}U^*.$$

推论4. 2. 3 若 A 为 Hermite 幂等阵（即正交投影阵），则 $A^+=A$.

注意到 Hermite 阵是正规阵，幂等阵的特征值只能是0和1，故从推论4. 2. 2立即得到本推论.

从定理4. 2. 4和定理4. 2. 5，我们知道，一般说来 A 与 A^+ 是不可交换的，即 $AA^+\neq A^+A$，但对正规阵却不然. 这就是下面的定理：

定理4. 2. 7 若 A 为正规阵，则

(a) $AA^+=A^+A$.

(b) 对任一自然数 k，$(A^k)^+=(A^-)^k$.

证明 利用推论4. 2. 2，证明是简单的.

注 注意对一般的矩阵 A，(b)不成立.

定理4. 2. 8 (a) $B^+A^+=\mathbf{0}\Longleftrightarrow AB=\mathbf{0}$.

(b) $AB^+=0 \Longleftrightarrow AB^*=0$.

(c) $B^+A=0 \Longleftrightarrow A^*B=0$.

证明 (a)充分性 若 $AB=0$,则 $B^*A^*=0$. 用 $(B^*B)^+$ 和 $(AA^*)^+$ 分别左乘和右乘此式,得

$$(B^*B)^+B^*A^*(AA^*)^+=0,$$

由定理4.2.2(d),上式即为 $B^+A^+=0$.

必要性 若 $B^+A^+=0$,则

$$\begin{aligned}0&=B^*BB^+A^+AA^*=B^*(BB^+)^*(A^+A)^*A^*\\&=B^*(B^+)^*B^*A^*(A^+)^*A^*\\&=B^*A^*=(AB)^*,\end{aligned}$$

于是 $AB=0$.

用类似的方法可以证明(b)和(c). 定理证毕.

§4.3 乘法公式

我们知道,若同阶方阵 A 与 B 皆可逆,则乘积矩阵 AB 也可逆,且乘法公式 $(AB)^{-1}=B^{-1}A^{-1}$ 成立. 但对 Moore-Penrose 逆,乘法公式

$$(AB)^+=B^+A^+ \tag{4.3.1}$$

一般说来不成立. 本节的目的是建立(4.3.1)成立的一些充要条件. 先证明一个引理.

引理4.3.1 设 $A\in C^{m\times n}$, $X\in C^{n\times m}$, $r(X)\leqslant r(A)$,

$$XA=A^+A, \quad AX=AA^+, \tag{4.3.2}$$

则必有 $X=A^+$.

证明 从(4.3.2)容易看出,我们只需证明 X 满足 $XAX=X$. 因为

$$r(X)\leqslant r(A)=r(AXA)\leqslant r(XA)\leqslant r(X),$$

于是

$$r(X)=r(XA),$$
$$\mathscr{M}(X)=\mathscr{M}(XA),$$

故存在矩阵 Y，使得 $X=XAY$，再用 XA 左乘得

$$XAX = XAXAY = XAY = X.$$

引理证毕.

定理4.3.1 乘法公式(4.3.1)成立，当且仅当下列两条件同时成立

$$A^+ABB^*A^* = BB^*A^*, \qquad (4.3.3)$$
$$BB^+A^*AB = A^*AB. \qquad (4.3.4)$$

证明 充分性 假设(4.3.3)和(4.3.4)都成立. 用 B^+ 和 $(AB)^{*+}$ 分别左、右乘(4.3.3)两边，其左边变为

$$B^+A^+ABB^*A^*(AB)^{*+} = B^+A^+(AB)(AB)^*(AB)^{*+}. \qquad (4.3.5)$$

应用定理4.2.1(c)和定理4.2.2(a)，得

$$AB(AB)^*(AB)^{*+} = AB.$$

于是，我们证明了

$$B^+A^+ABB^*A^*(AB)^{*+} = B^+A^+AB. \qquad (4.3.6)$$

再看(4.3.3)的右边，

$$B^+BB^*A^*(AB)^{*+} = (B^+BB^*)A^*(AB)^{*+}$$
$$=B^*A^*(AB)^{*+} = (AB)^*(AB^*)^+, \qquad (4.3.7)$$

这里我们利用了 $B^+BB^*=P_B\cdot B^*=B^*$. 结合(4.3.6)，我们证明了

$$B^+A^+AB=(AB)^*(AB)^{*+}=((AB)^+AB)^*$$
$$=(AB)^-AB. \qquad (4.3.8)$$

现在考虑(4.3.4). 将其两边做转置共轭运算，得

$$B^*A^*A(B^+)^*B^* = B^*A^*A. \qquad (4.3.9)$$

用$(AB)^{*+}$和A^+分别左乘、右乘上式两边，左边变为

$$\begin{aligned}&(AB)^{*+}B^*A^*AB^{*+}B^*A^+\\&=(AB)^{*+}(AB)^*AB^{*+}B^*A^+\\&=(AB)^{*+}(AB)^*A(BB^+)^*A^+\\&=(AB)^{*+}(AB)^*ABB^+A^+\\&=ABB^+A^+,\end{aligned}\tag{4.3.10}$$

这里我们应用了关系

$$(AB)^{*+}(AB)^*AB=(AB)^{*+}(AB)^*((AB)^*)^*=AB.$$

而(4.3.9)的右边在分别左乘和右乘$(AB)^{*+}$和A^+之后，变为

$$\begin{aligned}&(AB)^{*+}B^*A^*AA^+=(AB)^{*+}B^*A^*\\&=(AB)^{*+}(AB)^*=(AB(AB)^+)^*\\&=AB(AB)^+,\end{aligned}\tag{4.3.11}$$

这里应用了

$$A^*AA^+=A^*P_A=A^*.$$

综合(4.3.10)和(4.3.11)，我们得到

$$ABB^+A^+=AB(AB)^+.\tag{4.3.12}$$

另外，因为

$$\begin{aligned}B^+A^+&=B^+BB^+A^+AA^+\\&=B^+(BB^+)^*(A^+A)^*A^+\\&=B^+B^{+*}B^*A^*A^{+*}A^+\\&=B^+B^{+*}(AB)^*A^{+*}A^+,\end{aligned}$$

于是

$$r(B^+A^+)\leqslant r(AB).$$

将上式与(4.3.8)、(4.3.12)结合在一起，再利用引理4.3.1便证明了(4.3.1).

必要性　因为

$$(AB)^{+}AB = P_{\mathcal{M}((AB)^{*})},$$

故

$$(AB)^{+}(AB)(AB)^{*} = (AB)^{*}.$$

若(4.3.1)成立，上式变为

$$B^{*}A^{*} = B^{+}A^{+}ABB^{*}A^{*}.$$

用 $ABB^{*}B$ 左乘上式并利用 $B^{*}BB^{+}=B^{*}$，得

$$\begin{aligned} ABB^{*}BB^{*}A^{*} &= ABB^{*}BB^{+}A^{+}ABB^{*}A^{*} \\ &= ABB^{*}A^{+}ABB^{*}A^{*}, \end{aligned}$$

此式蕴含着

$$ABB^{*}(I - A^{+}A)BB^{*}A^{*} = 0.$$

因为 $I-A^{+}A$ 为半正定 Hermite 阵，故从上式可推出

$$(I - A^{+}A)BB^{*}A^{*} = 0,$$

这就证明了(4.3.3).

完全类似地，可证明(4.3.4). 定理证毕.

推论4.3.1 乘法公式(4.3.1)成立，当且仅当

$$\mathcal{M}(BB^{*}A^{*}) \subset \mathcal{M}(A^{*}), \qquad (4.3.13)$$

$$\mathcal{M}(A^{*}AB) \subset \mathcal{M}(B). \qquad (4.3.14)$$

证明 我们只需证明(4.3.3)和(4.3.4)与(4.3.13)和(4.3.14)等价，因为

$$A^{+}A = P_{\mathcal{M}(A^{+}A)} = P_{\mathcal{M}(A^{*})},$$

$$BB^{+} = P_{\mathcal{M}(BB^{+})} = P_{\mathcal{M}(B)}.$$

于是

$$\begin{aligned} A^{+}ABB^{*}A^{*} &= P_{\mathcal{M}(A^{*})}BB^{*}A^{*} \\ &= BB^{*}A^{*} \Longleftrightarrow (4.3.13)\text{ 成立}, \end{aligned}$$

$$\begin{aligned} BB^{+}A^{*}AB &= P_{\mathcal{M}(B)}A^{*}AB \\ &= A^{*}AB \Longleftrightarrow (4.3.14)\text{ 成立}. \end{aligned}$$

推论4.3.2 若 A^+ABB^* 和 A^*ABB^+ 皆为 Hermite 阵，则乘法公式(4.3.1)成立.

证明 若 A^+ABB^* 为 Hermite 阵，则

$$A^+ABB^* = BB^*(A^+A)^*.$$

用 A^* 右乘上式，得

$$A^+ABB^*A^* = BB^*(A^+A)^*A^* = BB^*A^*.$$

即(4.3.3)成立.

若 A^*ABB^+ 为 Hermite 阵，则

$$A^*ABB^+ = (BB^+)^*A^*A = BB^+A^*A.$$

用 B 右乘上式，得

$$BB^+A^*AB = A^*ABB^+B = A^*AB.$$

这就证明了(4.3.4). 根据定理(4.3.1)，于是(4.3.1)成立. 定理证毕.

§4.4 $(A+bc^*)^+$

设 $A\in C^{m\times n}$，$b\in C^m$，$c\in C^n$，形如 $A+bc^*$ 的矩阵称为秩1修正阵(rank one modified matrix). 此名称的含义是明显的. 本节讨论这样的矩阵的 Moore-Penrose 逆的表达式.

记

$$u=(I-AA^+)b,\qquad v^*=c^*(I-A^+A),$$

$$w=1+c^*A^+b,\qquad g=A^+b,$$

$$h^*=c^*A^+.$$

引理4.4.1 $Av^{*+}=0$，$u^+A=0$，$u^+b=1$，$v^+c=1$，这里 $v\neq 0$.

证明 利用推论4.1.4，容易证明这些事实.

引理4.4.2

$$r(A+bc^*)=r\begin{pmatrix} A & u \\ v^* & -w \end{pmatrix}-1.$$

证明 因为

$$\begin{pmatrix} A+bc^* & b \\ 0 & -1 \end{pmatrix}=\begin{pmatrix} I & 0 \\ h^* & 1 \end{pmatrix}\begin{pmatrix} A & u \\ v^* & -w \end{pmatrix}\begin{pmatrix} I & g \\ 0 & 1 \end{pmatrix}\begin{pmatrix} I & 0 \\ c^* & 1 \end{pmatrix}.$$

于是

$$r(A+bc^*)+1=r\begin{pmatrix} A & u \\ v^* & -w \end{pmatrix}.$$

引理证毕.

定理4.4.1

(a) 若 $u\neq 0, v\neq 0$, 则

$$(A+bc^*)^+=A^+-gu^+-(hv^+)^*+wv^{*+}u^+. \tag{4.4.1}$$

(b) 若 $u=0, v\neq 0, w=0$, 则

$$(A+bc^*)^+=A^+-gg^+A^+-(v^*)^+h. \tag{4.4.2}$$

(c) 若 $u=0, w\neq 0$, 则

$$(A+bc^*)^+=A^++\frac{1}{\overline{w}}vg^*A^+-\frac{\overline{w}}{\sigma_1}p_1q_1^*, \tag{4.4.3}$$

其中

$$p_1=-\left(\frac{\|g\|^2}{\overline{w}}v+g\right),$$

$$q_1^*=-\left(\frac{\|v\|^2}{\overline{w}}g^*A^++h^*\right),$$

$$\sigma_1=\|g\|^2\|v\|^2+|w|^2.$$

(d) 若 $u\neq 0, v=0, w=0$, 则

$$(A+bc^*)^+=A^+-A^+hh^+-gu^+. \tag{4.4.4}$$

(e) 若 $v=0, w\neq 0$, 则

$$(A+bc^*)^+=A^++\frac{1}{w}A^+hu^*-\frac{\overline{w}}{\sigma_2}p_2q_2^*, \quad (4.4.5)$$

其中

$$p_2=-\left(\frac{\|u\|^2}{\overline{w}}A^+h+g\right),$$

$$q_2^*=-\left(\frac{\|h\|^2}{\overline{w}}u^*+h^*\right),$$

$$\sigma_2=\|h\|^2\|u\|^2+|w|^2.$$

(f) 若 $u=0, v=0, w=0$，则

$$(A+bc^*)^+=A^+-gg^+A^+-A^+hh^++g^+A^+(h^*)^+gh^*. \quad (4.4.6)$$

证明 记 $M=A+bc^*$.

(a) 用 X_1 表示(4.4.1)的右端，利用引理4.4.1以及 $c^*g=w-1, b-Ag=u, h^*b=w-1$ 和 $v^*=c^*-h^*A$，可以证明

$$MX_1=AA^++uu^+,$$

$$X_1M=A^+A+vv^+.$$

据此不难验证，X_1 满足 Penrose 方程.

(b) 记 X_2 为(4.4.2)的右端，类似于(a)，注意到对现在的情况，$w=0, u=0$，可以证得

$$MX_2=AA^+,$$

$$X_2M=A^+A-gg^++vv^+.$$

根据这两个关系，可以验证，X_2 满足 Penrose 方程.

(c) 记 X_3 为(4.4.3)的右端. 因为 $u=0$，故 $b\in\mathscr{M}(A)$，因而 $\mathscr{M}(M)\subset\mathscr{M}(A)$. 由 $w\neq0$ 及引理4.4.2知

$$r(M)=r(A), \quad (4.4.7)$$

从而有

$$\mathscr{M}(M)=\mathscr{M}(A).$$

依定理4.2.4有

$$MM^{+} = P_{M} = P_{A} = AA^{-}, \qquad (4.4.8)$$

但 $q_1^* AA^{+} = q_1^*$，因而

$$X_3MM^{-} = X_3AA^{+} = X_3. \qquad (4.4.9)$$

若能证明

$$M^{+} M = X_3M, \qquad (4.4.10)$$

则利用(4.4.9)有

$$X_3 = X_3MM^{+} = M^{+} MM^{+} = M^{+}.$$

于是，问题归结为证明(4.4.10).

为证(4.4.10)，我们证明

$$M^{+} M = A^{+} A - gg^{+} + p_1p_1^{+}, \qquad (4.4.11)$$

$$X_3M = A^{+} A - gg^{+} + p_1p_1^{+}. \qquad (4.4.12)$$

记(4.4.11)右端为 Q. 因为 $M^{+}M$ 为向 $\mathscr{M}(M^{+}) = \mathscr{M}(M^{*})$ 上的正交投影阵，故我们需要证明 Q 为向 $\mathscr{M}(M^{+})$ 上的正交投影阵.

首先，容易验证 Q 为幂等阵，因为它是 Hermite 阵，故它一定是正交投影阵. 于是

$$\begin{aligned} r(Q) &= \mathrm{tr}(Q) = \mathrm{tr}(A^{+} A) - \mathrm{tr}(gg^{+}) + \mathrm{tr}(p_1p_1^{+}) \\ &= \mathrm{tr}(A^{+} A) - r(gg^{+}) + r(p_1p_1^{+}) \\ &= \mathrm{tr}(A^{+} A) = r(A^{+} A), \end{aligned}$$

这里利用了 gg^{+} 和 $p_1p_1^{+}$ 都是秩为1的幂等阵以及 $A^{+}A$ 为幂等阵. 再利用(4.4.7)，得

$$r(Q) = r(A^{+} A) = r(A) = r(M). \qquad (4.4.13)$$

但直接计算可知

$$MQ = M \Rightarrow QM^{*} = M^{*},$$

因而

$$\mathscr{M}(M^*) \subset \mathscr{M}(Q). \tag{4.4.14}$$

结合(4.4.13)得:$\mathscr{M}(M^*)=\mathscr{M}(Q)$.因而 Q 为向 $\mathscr{M}(M^*)=\mathscr{M}(M^+)$上的正交投影阵.这就证明了(4.4.11).

再证(4.4.12).因为 $A^+Ag=g$,

$$q_1^* b = 1 - \sigma_1 \overline{w}^{-1},$$

$$q_1^* A + c^* = \frac{-\|v\|^2}{\overline{w}} g^* + v^*,$$

于是

$$\begin{aligned} X_3 M &= A^+ A + \frac{1}{\overline{w}} v g^* - \frac{\overline{w}}{\sigma_1} p_1 q_1^* A - p_1 c^* - \frac{\overline{w}}{\sigma_1} p_1 c^* \\ &= A^+ A + \frac{1}{\overline{w}} v g^* - \frac{\overline{w}}{\sigma_1} p_1 (q_1^* A + c^*) \\ &= A^+ A + \frac{1}{\overline{w}} v g^* - \frac{\overline{w}}{\sigma_1} p_1 \left(v - \frac{\|v\|^2}{\overline{w}} g^* \right). \end{aligned}$$

再利用

$$v = - w\|g\|^{-2}(p_1^* + g),$$

$$\frac{1}{\|p_2\|^2} = \frac{|w|^2}{\sigma_1}\|g\|^2,$$

得到

$$\begin{aligned} X_3 M &= A^+ A + \frac{1}{\overline{w}} v g^* + \frac{1}{\|g\|^2} p_1 g^* + p_1 p_1^+ \\ &= A^+ A - g g^+ + p_1 p_1^+, \end{aligned}$$

在最后一等式,我们利用了

$$\frac{1}{\overline{w}} v + \frac{1}{\|g\|^2} p_1 = -\frac{1}{\|g\|^2} g,$$

$$g g^+ = \frac{1}{\|g\|^2} g g^*.$$

(4.4.12)得证.

(d)和(e)的证明：注意到

$$(\boldsymbol{M}^*)^+ = (\boldsymbol{M}^+)^*,$$

而从

$$\boldsymbol{M}^* = \boldsymbol{A}^* + \boldsymbol{c}\boldsymbol{b}^*$$

以及 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 的定义可知，对 $\boldsymbol{M}^*$，(b)的假设成立. 于是对 $\boldsymbol{M}^*$ 可应用(4.4.2)，然后两边取转置和共轭即得(4.4.4). 类似地，利用(c)可证明(e).

(f) 我们先证明

$$\boldsymbol{M}\boldsymbol{M}^+ = \boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+, \qquad (4.4.15)$$

$$\boldsymbol{M}^+\boldsymbol{M} = \boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{g}\boldsymbol{g}^+. \qquad (4.4.16)$$

不难证明，$\boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+$ 和 $\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{g}\boldsymbol{g}^+$ 都是 Hermite 幂等阵，因此都是正交投影阵，故

$$\begin{aligned} r(\boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+) &= \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+) \\ &= \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^+) - \mathrm{tr}(\boldsymbol{h}\boldsymbol{h}^+) \\ &= r(\boldsymbol{A}\boldsymbol{A}^+) - r(\boldsymbol{h}\boldsymbol{h}^+) \\ &= r(\boldsymbol{A}) - 1. \end{aligned}$$

同理可证得

$$r(\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{g}\boldsymbol{g}^+) = r(\boldsymbol{A}) - 1.$$

对现在的情况：$\boldsymbol{u}=\boldsymbol{0}$，$\boldsymbol{v}=\boldsymbol{0}$，$w=0$，于是从引理4.4.2可得

$$r(\boldsymbol{M}) = r(\boldsymbol{A}) - 1.$$

因而我们证明了

$$r(\boldsymbol{M}) = r(\boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+) = r(\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{g}\boldsymbol{g}^+). \qquad (4.4.17)$$

直接计算可验证

$$(\boldsymbol{A}\boldsymbol{A}^+ - \boldsymbol{h}\boldsymbol{h}^+)\boldsymbol{M} = \boldsymbol{M},$$

$$\boldsymbol{M}(\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{g}\boldsymbol{g}^+) = \boldsymbol{M}.$$

故

$$\mathscr{M}(M) \subset \mathscr{M}(AA^+ - hh^+),$$
$$\mathscr{M}(M^*) \subset \mathscr{M}(A^+A - gg^+).$$

结合(4.4.17),我们得到

$$\mathscr{M}(M) = \mathscr{M}(AA^+ - hh^+),$$
$$\mathscr{M}(M^*) = \mathscr{M}(A^+A - gg^+).$$

从而有

$$MM^+ = P_M = AA^+ - hh^+,$$
$$M^+M = P_{M^*} = A^+A - gg^+,$$

这就证明了(4.4.15)和(4.4.16).

用 X_4记(4.4.6)的右端,利用(4.4.15)和(4.4.16)分别得到

$$X_4MM^+ = X_4(AA^+ - hh^+) = X_4,$$
$$X_4M = A^+A - gg^+ = M^+M,$$

于是

$$M^+ = (M^+M)M^+ = X_4MM^+ = X_4.$$

定理证毕.

推论4.4.1 若 $b \in \mathscr{M}(A)$,$c \in \mathscr{M}(A^*)$,则

$$(A + bc^*)^+ = A^+ - \frac{1}{w}A^+bc^*A^+,$$

其中 $w = 1 + c^*A^+b \neq 0$.

证明 因为 $b \in \mathscr{M}(A) \Longleftrightarrow u = 0$,而 $c \in \mathscr{M}(A^*) \Longleftrightarrow v = 0$,应用定理中(c)和(e),可得所要结论.

§4.5 正交投影阵与线性流形

本节的目的有两个,其一是利用 Moore-Penrose 广义逆对正交投影阵做一些更深入讨论,给出两个子空间的和空间

与交空间上的正交投影阵的表示；其二是用正交投影阵研究线性流形.

设 L 和 M 为 C^n 的子空间，P_L 和 P_M 分别表示向 L 和 M 的正交投影阵.

引理4.5.1

$$(P_L \vdots \pm P_M)^+ = \begin{pmatrix} P_L \\ \pm P_M \end{pmatrix} (P_L + P_M)^+.$$

证明 利用定理4.2.2(d)以及 P_L 和 P_M 皆为 Hermite 幂等阵，结论容易得证.

定理4.5.1

$$P_{L+M} = (P_L + P_M)(P_L + P_M)^+ = (P_L + P_M)^+ (P_L + P_M).$$

证明 因为 $L+M=\mathscr{M}(P_L \vdots P_M)$. 利用(4.2.2)和引理4.5.1得

$$P_{L+M} = P_{(P_L \vdots P_M)} = (P_L \vdots P_M)(P_L \vdots P_M)^+$$

$$= (P_L \vdots P_M)\begin{pmatrix} P_L \\ P_M \end{pmatrix}(P_L + P_M)^+.$$

因为 P_L 和 P_M 皆为 Hermite 幂等阵，故

$$P_{L+M} = (P_L + P_M)(P_L + P_M)^+,$$

第一个等式得证. 再由 P_L+P_M 是 Hermite 阵以及对 Hermite 阵 A，有 $AA^+=A^+A$，第二个等式得证.

定理4.5.2

$$P_{L\cap M} = 2P_L(P_L + P_M)^+ P_M = 2P_M(P_L + P_M)^+ P_L. \tag{4.5.1}$$

证明 因为 $M\subset L+M$，故有

$$P_{L+M}P_M = P_M = P_M P_{L+M}.$$

利用定理4.5.1，

$$(P_L + P_M)(P_L + P_M)^+ P_M$$
$$= P_M = P_M(P_L + P_M)^+ (P_L + P_M).$$
从上式的左端和右端分别减去 $P_M(P_L+P_M)^+P_M$ 得
$$P_L(P_L + P_M)^+ P_M = P_M(P_L + P_M)^+ P_L,$$
这就证明了(4.5.1)的第二个等号. 记(4.5.1)右边的矩阵为 H,则有 $\mathscr{M}(H)\subset L\cap M$,于是
$$\begin{aligned} H &= P_{L\cap M}H \\ &= P_{L\cap M}(P_L(P_L + P_M)^+ P_M + P_M(P_L + P_M)^+ P_L) \\ &= P_{L\cap M}P_L(P_L + P_M)^+ P_M + P_{L\cap M}P_M(P_L + P_M)^+ P_L. \end{aligned}$$
利用 $P_{L\cap M}P_L=P_{L\cap M}$,$P_{L\cap M}P_M=P_{L\cap M}$以及定理4.5.1,得
$$\begin{aligned} H &= P_{L\cap M}(P_L + P_M)^+ (P_L + P_M) \\ &= P_{L\cap M}P_{L+M} \\ &= P_{L\cap M}, \end{aligned}$$
最后的等式是因为 $L\cap M\subset L+M$. 定理证毕.

现在讨论第二个问题,先引进线性流形(Linear manifold)的概念.

设 $x_0\in C^n$,S 为 C^n 的一个子空间,我们称集合
$$x_0 + S = \{x_0 + y : y \in S\}$$
为 C^n 中的一个线性流形,简称流形. 若 $x_0\in S$,则 $x_0+S=S$,因此,子空间也是流形,但流形不必是子空间.

容易证明,两个流形之和仍为流形,即 $x_0,y_0\in C^n$,S_1和 S_2 为 C^n 的两个子空间,则$\{x_0+S_1\}+\{y_0+S_2\}$也是流形. 若两个流形的交是非空的,则它必是一个流形. 这个事实的证明较麻烦,我们先证明一个引理.

设 A 为 $m\times n$ 阵,S 为 C^n 的一个子空间,记
$$AS = \{x : x = Ay, y \in S\}.$$

它是 C^m 的子空间.

引理4.5.2 设 L 和 M 为 C^n 的两个子空间,则

(a) $L\cap M=(P_L \vdots 0)\mathcal{N}(P_L \vdots -P_M)$

$\qquad\qquad =(0 \vdots P_M)\mathcal{N}(P_L \vdots -P_M)$.

(b) $L\cap M=\mathcal{N}(P_{L^\perp}+P_{M^\perp})$.

(c) $L\cap M=\mathcal{N}(I-P_LP_M)=\mathcal{N}(I-P_MP_L)$.

证明 (a)显然,$x\in L\cap M \Longleftrightarrow \exists\ u,v\in C^n$,使得

$$x=P_Lu=P_Mv$$

$$\Longleftrightarrow x=P_Lu\ ,(P_L \vdots -P_M)\begin{pmatrix}u\\v\end{pmatrix}=0$$

或

$$x=P_Mv\ ,(P_L \vdots -P_M)\begin{pmatrix}u\\v\end{pmatrix}=0$$

$$\Longleftrightarrow x=(P_L \vdots 0)\begin{pmatrix}u\\v\end{pmatrix},\quad \begin{pmatrix}u\\v\end{pmatrix}\in\mathcal{N}(P_L \vdots -P_M)$$

或

$$x=(0 \vdots P_M)\begin{pmatrix}u\\v\end{pmatrix},\quad \begin{pmatrix}u\\v\end{pmatrix}\in\mathcal{N}(P_L \vdots -P_M)$$

$$\Longleftrightarrow L\cap M=(P_L \vdots 0)\mathcal{N}(P_L \vdots -P_M)$$

$$=(0 \vdots P_M)\mathcal{N}(P_L \vdots -P_M).$$

(b) 若 $x\in L\cap M$,则

$$P_{L^\perp}x=P_{M^\perp}x=0,$$

即

$$x\in\mathcal{N}(P_{L^\perp}+P_{M^\perp}).$$

反过来,若

$$x\in\mathcal{N}(P_{L^\perp}+P_{M^\perp}),$$

则

$$P_{L^\perp}x+P_{M^\perp}x=0,$$

即

$$(\boldsymbol{I}-\boldsymbol{P}_L)\boldsymbol{x}+(\boldsymbol{I}-\boldsymbol{P}_M)\boldsymbol{x}=\boldsymbol{0},$$

也就是

$$2\boldsymbol{x}=\boldsymbol{P}_L\boldsymbol{x}+\boldsymbol{P}_M\boldsymbol{x}.$$

两边取范数，利用三角不等式及$\|\boldsymbol{P}_L\boldsymbol{x}\|\leqslant\|\boldsymbol{x}\|$，$\|\boldsymbol{P}_M\boldsymbol{x}\|\leqslant\|\boldsymbol{x}\|$，得

$$2\|\boldsymbol{x}\|\leqslant\|\boldsymbol{P}_L\boldsymbol{x}\|+\|\boldsymbol{P}_M\boldsymbol{x}\|\leqslant 2\|\boldsymbol{x}\|.$$

于是

$$\|\boldsymbol{P}_L\boldsymbol{x}\|=\|\boldsymbol{x}\|=\|\boldsymbol{P}_M\boldsymbol{x}\|,$$

这就证明了

$$\boldsymbol{P}_L\boldsymbol{x}=\boldsymbol{x}=\boldsymbol{P}_M\boldsymbol{x},$$

因此 $\boldsymbol{x}\in L\cap M$.

(c) 设 $\boldsymbol{x}\in L\cap M$，则

$$\boldsymbol{x}=\boldsymbol{P}_L\boldsymbol{x}=\boldsymbol{P}_M\boldsymbol{x}=\boldsymbol{P}_L\boldsymbol{P}_M\boldsymbol{x}.$$

因而

$$\boldsymbol{x}\in\mathscr{N}(\boldsymbol{I}-\boldsymbol{P}_L\boldsymbol{P}_M).$$

反之，若

$$\boldsymbol{x}\in\mathscr{N}(\boldsymbol{I}-\boldsymbol{P}_L\boldsymbol{P}_M),$$

则有

$$\boldsymbol{x}=\boldsymbol{P}_L\boldsymbol{P}_M\boldsymbol{x}\in L. \tag{4.5.2}$$

另一方面，因为

$$\boldsymbol{x}=\boldsymbol{P}_M\boldsymbol{x}+\boldsymbol{P}_{M^\perp}\boldsymbol{x},\quad \boldsymbol{P}_M\boldsymbol{x}\perp\boldsymbol{P}_{M^\perp}\boldsymbol{x},$$

所以

$$\|\boldsymbol{P}_M\boldsymbol{x}\|^2+\|\boldsymbol{P}_{M^\perp}\boldsymbol{x}\|^2=\|\boldsymbol{x}\|^2=\|\boldsymbol{P}_L\boldsymbol{P}_M\boldsymbol{x}\|^2\leqslant\|\boldsymbol{P}_M\boldsymbol{x}\|^2,$$

于是

$$\boldsymbol{P}_{M^\perp}\boldsymbol{x}=\boldsymbol{0},$$

这就表明 $\boldsymbol{x}\in M$，因而 $\boldsymbol{x}=\boldsymbol{P}_M\boldsymbol{x}$. 从(4.5.2)我们有

$$\boldsymbol{x}=\boldsymbol{P}_L\boldsymbol{x}\in L,$$

于是第一个等式得证. 由 L 和 M 的对称性知第二个等式也成立. 引理证毕.

定理4.5.3 设 x 和 y 为 C^n 中的两个向量, L 和 M 为 C^n 中的两个子空间, 则两个流形的交

$$\{x+L\}\cap\{y+M\} \tag{4.5.3}$$

非空 $\Longleftrightarrow$

$$x-y\in L+M. \tag{4.5.4}$$

若(4.5.4)成立, 则

(a) $\{x+L\}\cap\{y+M\}=x+P_L(P_L+P_M)^+(y-x)+(L\cap M)$.

(b) $\{x+L\}\cap\{y+M\}=x+(P_{L^\perp}+P_{M^\perp})^+P_{M^\perp}(y-x)+(L\cap M)$.

(c) $\{x+L\}\cap\{y+M\}=x+(I-P_MP_L)^+P_{M^\perp}(y-x)+(L\cap M)$.

证明 集合(4.5.3)非空, 当且仅当存在 $u\in L, v\in M$, 使得

$$x+u=y+v,$$

即

$$x-y=-u+v\in L+M.$$

(a) $z\in\{x+L\}\cap\{y+M\}\Longleftrightarrow\exists\ u,v\in C^n$, 使得

$$z=x+P_Lu=y+P_Mv \tag{4.5.5}$$

$$\Longleftrightarrow(P_L\ \vdots\ -P_M)\begin{pmatrix}u\\v\end{pmatrix}=y-x.$$

因为(4.5.4)成立, 故此方程组是相容的, 于是, 它的通解为

$$\begin{pmatrix}u\\v\end{pmatrix}=(P_L\ \vdots\ -P_M)^+(y-x)$$

$$+(I-(P_L\ \vdots\ -P_M)^+(P_L\ \vdots\ -P_M))t$$

$$= \begin{pmatrix} P_L \\ -P_M \end{pmatrix} (P_L + P_M)^+ (y - x)$$

$$+ (I - (P_L \vdots -P_M)^- (P_L \vdots -P_M))t.$$

代入(4.5.5),得

$$z = x + (P_L \vdots 0) \begin{pmatrix} u \\ v \end{pmatrix}$$

$$= x + P_L(P_L + P_M)^+ (y - x)$$

$$+ (P_L \vdots 0)(I - (P_L \vdots -P_M)^- (P_L \vdots -P_M))t.$$

结合引理4.5.2(a),我们证明了

$$z \in \{x + L\} \cap \{y + M\}$$

$$\Longleftrightarrow z = x + P_L(P_L + P_M)^+ (y - x) + (L \cap M).$$

(b) 将(4.5.5)改写为

$$P_L u - P_M v = y - x.$$

用 $P_{M^\perp}$ 左乘上式两端,得

$$P_{M^\perp} P_L u = P_{M^\perp}(y - x), \qquad (4.5.6)$$

上式等价于

$$(P_{L^\perp} + P_{M^\perp}) P_L u = P_{M^\perp}(y - x).$$

显然,这个方程组是相容的,其通解可表为

$$P_L u = (P_{L^\perp} + P_{M^\perp})^+ P_{M^\perp}(y - x)$$

$$+ (I - (P_{L^\perp} + P_{M^\perp})^- (P_{L^\perp} + P_{M^\perp}))t,$$

也可写为

$$P_L u = (P_{L^\perp} + P_{M^\perp})^+ P_{M^\perp}(y - x) + \mathcal{N}(P_{L^\perp} + P_{M^\perp}).$$

利用引理4.5.2(b),得

$$P_L u = (P_{L^\perp} + P_{M^\perp})^+ P_{M^\perp}(y - x) + (L \cap M),$$

代入(4.5.5),(b)得证.

(c) 将方程(4.5.6)改写为

$$(I - P_M P_L) P_L u = P_{M^\perp}(y - x),$$

显然,它的通解可表为

$$P_L u = (I - P_M P_L)^+ P_{M^\perp}(y - x) + \mathcal{N}(I - P_M P_L)$$
$$= (I - P_M P_L)^+ P_{M^\perp}(y - x) + L \cap M.$$

这里应用了引理4.5.2(c)(代入4.5.5).(c)得证.定理证毕.

根据 x,y 和 L,M 在问题中的对称性,由定理4.5.3不难获得如下推论:

推论4.5.1 在定理4.5.3的假设下,若(4.5.4)成立,则

(a) $\{x+L\}\cap\{y+M\}=y-P_M(P_L+P_M)^+(y-x)+(L\cap M)$.

(b) $\{x+L\}\cap\{y+M\}=y-(P_{L^\perp}+P_{M^\perp})^+P_{L^\perp}(y-x)+(L\cap M)$.

(c) $\{x+L\}\cap\{y+M\}=y-(I-P_LP_M)^+P_{L^\perp}(y-x)+(L\cap M)$.

定理4.5.3和推论4.5.1表明,若条件(4.5.4)成立,则两个流形的交也是流形,我们把这个事实归纳成如下定理:

定理4.5.4 设 x,y 为 C^n 中的两个向量,L 和 M 为 C^n 中的两个子空间,若 $x-y\in L+M$,则

$$\{x + L\} \cap \{y + M\}$$

为一流形.

在流形 $x+L$ 中,x 不是唯一的.即$\exists$ $y\neq x$ 使得

$$x + L = y + L.$$

事实上,$\forall$ $u\in L$,取 $y=x+u$,上式仍成立.注意到

$$x + L = (x - P_L x) + L = P_{L^\perp} x + L, \quad (4.5.7)$$

而 $x_0=P_{L^\perp}x\perp L$,故流形 $x+L$ 的这个表示称为正交表示.

推论4.5.2 设 L 和 M 是 C^n 中的两个子空间,且

$$x \in L^\perp, \quad y \in M^\perp. \quad (4.5.8)$$

若$\{x+L\}\cap\{y+M\}$非空，则定理4.5.3和推论4.5.1给出的$\{x+L\}\cap\{y+M\}$的前两种表示都是正交表示.

证明 这四种表示都具有共同形式

$$\{x+L\}\cap\{y+M\}=z+(L\cap M). \tag{4.5.9}$$

因此，要证明(4.5.9)是正交表示，只需证明

$$P_{L\cap M}z=0. \tag{4.5.10}$$

定理4.5.3(a)的正交性证明：对这种情形

$$z=x+P_L(P_L+P_M)^+(y-x).$$

因为$(L\cap M)^{\perp}\subset L^{\perp}+M^{\perp}$，故由(4.5.8)知

$$P_{L\cap M}x=0,\quad P_{L\cap M}y=0. \tag{4.5.11}$$

利用

$$P_{L\cap M}(P_L+P_M)^+=(P_L+P_M)^+P_{L\cap M}$$

和

$$P_{L\cap M}=P_LP_{L\cap M}=P_{L\cap M}P_L=P_MP_{L\cap M}=P_{L\cap M}P_M$$

以及(4.5.11)，我们得到

$$\begin{aligned}P_{L\cap M}z&=P_{L\cap M}x+P_{L\cap M}P_L(P_L+P_M)^+(y-x)\\&=P_{L\cap M}x+(P_L+P_M)^+P_{L\cap M}(y-x)\\&=0.\end{aligned}$$

这就证明了定理4.5.3(a)给出的是正交表示.

推论4.5.1(a)的正交性证明：用完全类似的方法可证明这一结论.

定理4.5.3(b)的正交性证明：对这种情形，

$$z=x+(P_{L^{\perp}}+P_{M^{\perp}})^+P_{M^{\perp}}(y-x).$$

因为

$$P_{L\cap M}(P_{L^{\perp}}+P_{M^{\perp}})^+=0,$$

结合(4.5.11)，我们有

$$P_{L\cap M}z = P_{L\cap M}x + P_{L\cap M}(P_{L^\perp} + P_{M^\perp})^+ P_{M^\perp}(y - x) = 0.$$

推论4.5.1(b)的正交性证明：若(4.5.8)成立，利用定理4.5.1得，

$$\begin{aligned} y - x &= P_{L^\perp + M^\perp}(y - x) \\ &= (P_{L^\perp} + P_{M^\perp})^+ (P_{L^\perp} + P_{M^\perp})(y - x), \end{aligned}$$

整理即得

$$\begin{aligned} & x + (P_{L^\perp} + P_{M^\perp})^+ P_{M^\perp}(y - x) \\ = & y - (P_{L^\perp} + P_{M^\perp})^+ P_{L^\perp}(y - x). \end{aligned}$$

推论证毕.

§4.6 展 开 定 理

设 A 为任意矩阵.本节要讨论的 A^+ 的展开定理,就是通过只包含A^*或 A^*A 的有限项乘积或无穷级数来表示 A^+,这包括 Lagrange-Sylvester 公式,Neumann 型展开和 Neumann-Euler 展开.

1. Lagrange 型展开

设 $f_k(x)$为不高于 k 次的多项式:

$$f_k(x) = a_0 + a_1x + \cdots + a_kx^k.$$

对 n 阶方阵 A,定义

$$f_k(A) = a_0I + a_1A + \cdots + a_kA^k.$$

若 A 相似于对角阵,即存在可逆阵 Q,使得

$$A = Q\Lambda Q^{-1},$$

其中 $\Lambda=\mathrm{diag}(\lambda_1 I_{r_1},\cdots,\lambda_t I_{r_t},0,\cdots,0)$,$\lambda_1,\cdots,\lambda_t$ 为 A 的互不相同的非零特征值,重数分别为 $r_1,\cdots,r_t$,$r=r(A)=\sum_{i=1}^{t} r_i$. 则

$$f_k(A) = Qf_k(\Lambda)Q^{-1}.$$

注意到 A^*A 的全部非零特征值为 $\lambda_1^2,\cdots,\lambda_t^2$，依定理2.7.1，存在 Hermite 幂等阵 $P_i,i=1,\cdots,t$，使得

$$A^*A=\sum_{i=1}^{t}\lambda_i^2P_i, \tag{4.6.1}$$

这里 P_i 满足

$$P_iP_j=0,\quad i\neq j, \tag{4.6.2}$$

$$I=\sum_{i=1}^{t}P_i, \tag{4.6.3}$$

于是

$$f_k(A^*A)=\sum_{i=1}^{t}f_k(\lambda_i^2)P_i.$$

记

$$\varphi(\lambda)=\prod_{i=1}^{t}(\lambda-\lambda_i^2),$$

$$\varphi_i(\lambda)=\prod_{j\neq i}(\lambda-\lambda_j^2),$$

利用(4.6.2)和(4.6.3)，不难证明

$$\varphi_i(A^*A)=\varphi_i(\lambda_i^2)P_i,\quad i=1,\cdots,t.$$

于是

$$P_i=\frac{\varphi_i(A^*A)}{\varphi_i(\lambda_i^2)},\quad i=1,\cdots,t.$$

将此式代入 $(A^*A)^+$ 的谱分解

$$(A^*A)^+=\sum_{i=1}^{t}\lambda_i^{-2}P_i$$

中，得到

$$(A^*A)^+=\sum_{i=1}^{t}\lambda_i^{-2}\frac{\prod_{j\neq i}(A^*A-\lambda_j^2I)}{\prod_{j\neq i}(\lambda_i^2-\lambda_j^2)}.$$

最后有

$$A^{+} = (A^{*}A)^{+}A^{*} = \sum_{i=1}^{t}\lambda_i^{-2}\frac{\prod_{j\neq i}(A^{*}A-\lambda_j^2 I)}{\prod_{j\neq i}(\lambda_i^2-\lambda_j^2)}A^{*}. \tag{4.6.4}$$

这就是 A 的 Lagrange 型展开，也称为 Lagrange-Sylvester 公式.

从前面的讨论知，展开式(4.6.1)并不要求 A 是方阵，于是我们证明了如下定理.

定理4.6.1 设 A 为任一矩阵，$\lambda_1^2,\cdots,\lambda_t^2$ 为 $A^{*}A$ 的全部非零特征值. 则 A^{+} 有展开式(4.6.4).

2. Neumann 型展开

设 $A\in C^{m\times n}$ 有如下奇异值分解：

$$A = P\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix}Q^{*}, \tag{4.6.5}$$

其中 P 和 Q 分别为 $m\times m$ 和 $n\times n$ 酉阵，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_t)$，$\lambda_i>0, i=1,\cdots,t$ 为 A 的奇异值. 记

$$c = \max(\lambda_1^2,\cdots,\lambda_t^2).$$

定理4.6.2 设$0<\alpha<2/c$，则

$$A^{+} = \alpha\sum_{k=0}^{\infty}(I_n-\alpha A^{*}A)^k A^{*}. \tag{4.6.6}$$

证明 由(4.6.5)得

$$A^{*}A = Q\begin{pmatrix}\Lambda^2 & 0\\ 0 & 0\end{pmatrix}Q^{*},$$

于是

$$(I_n-\alpha A^{*}A)^k = Q\left(I_n-\alpha\begin{pmatrix}\Lambda^2 & 0\\ 0 & 0\end{pmatrix}\right)^k Q^{*},$$

$$\alpha(I_n - \alpha A^* A)^k A^* = \alpha Q\left(I_n - \alpha\begin{pmatrix} \Lambda^2 & 0 \\ 0 & 0 \end{pmatrix}\right)^k \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} P^*$$

$$= Q D_k(\alpha) P^*, \tag{4.6.7}$$

其中

$$D_k(\alpha) = \mathrm{diag}(d_1^{(k)}, \cdots, d_n^{(k)}), \tag{4.6.8}$$

$$d_i^{(k)} = \begin{cases} \alpha(1 - \alpha\lambda_i^2)^k \lambda_i, & i = 1, \cdots, t, \\ 0, & i > t. \end{cases}$$

于是

$$\sum_{k=0}^{\infty} d_i^{(k)} = \alpha\lambda_i \sum_{k=0}^{\infty} (1 - \alpha\lambda_i^2)^k = \frac{1}{\lambda_i}, \quad i = 1, \cdots, t. \tag{4.6.9}$$

为了此级数收敛，我们要求

$$\alpha < \frac{2}{\lambda_i}, \quad i = 1, \cdots, t,$$

这等价于 $\alpha < 2/c$. 于是从(4.6.7)和(4.6.9)我们有

$$\alpha \sum_{k=0}^{\infty} (I_n - \alpha A^* A)^k A^*$$

$$= Q\left[\sum_{k=0}^{\infty} D_k(\alpha)\right] P^* = Q\begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} P^* = A^+.$$

定理证毕.

从定理证明过程容易推出如下事实.

推论4.6.1 设 $0 < \alpha < 2/c$，则

$$A^+ = \alpha \sum_{k=0}^{\infty} A^* (I_m - \alpha A A^*)^k.$$

推论4.6.2 若 A 可逆，$0 < \alpha < \frac{2}{c}$. 则

$$A^{-1} = \alpha \sum_{k=0}^{\infty} A^* (I_m - \alpha A A^*)^k$$

$$=\alpha\sum_{k=0}^{\infty}(I_n-\alpha A^*A)^kA^*.$$

用类似的方法可以证明

定理4.6.3 设 $A\in C^{m\times n}$,则

$$A^+=\sum_{k=1}^{\infty}A^*(I_m+AA^*)^{-k}$$

$$=\sum_{k=1}^{\infty}(I_n+A^*A)^{-k}A^*.$$

3. Neumann-Euler 展开

定理4.6.4 设 $A\in C^{m\times n}$,$0<\alpha<\frac{2}{c}$. 记

$$A_p^+=\alpha\Big\{[I_n+(I_n-\alpha A^*A)]\times\prod_{k=1}^{p-1}[I_n+(I_n-\alpha A^*A)^{2^k}]\Big\}A^*,$$

则

$$\|A^+-A_p^+\|_2\leqslant\max_i\frac{(1-\alpha\lambda_i^2)^{2^p}}{\lambda_i},$$

这里$\|A\|_2$表示矩阵 A 的谱范数.

证明 利用 Euler 恒等式

$$(1+x)\prod_{k=1}^{p-1}(1+x^{2^k})=\sum_{k=0}^{2^p-1}x^k,$$

于是

$$[I_n+(I_n-\alpha A^*A)]\prod_{k=1}^{p-1}[I_n+(I_n-\alpha A^*A)^{2^k}]$$

$$=\sum_{k=0}^{2^p-1}(I_n-\alpha A^*A)^k.$$

结合定理4.6.2,我们有

$$A^+ - A_p^+ = \alpha \sum_{k=2^p}^{\infty} (I_n - \alpha A^* A)^k A^*$$

$$= \alpha Q \sum_{k=2^p}^{\infty} D_k(\alpha) P^* ,$$

这里 $D_k(\alpha)$的定义同(4.6.8). 由谱范数的相容性(见§2.9)得

$$\|A^+ - A_p^+\|_2 \leqslant \alpha \|Q\|_2 \| \sum_{k=2^p}^{\infty} D_k(\alpha) \|_2 \|P^*\|_2$$

$$= \max_i \frac{(1-\alpha\lambda_i^2)^{2^p}}{\lambda_i}.$$

定理证毕.

§4.7 连续性问题

设 $A \in C^{m\times n}, E \in C^{m\times n}$. 将 E 视为对 A 的某种扰动或摄动. 若

$$\lim_{E\to 0} (A+E)^+ = A^+ \tag{4.7.1}$$

成立, 则 A^+ 是 A 的元素的连续函数, 简称 A^+ 是连续的. 但是, 一般说来 A^+ 并不是连续的. 下面是一个简单例子. 例如

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 \\ 0 & \varepsilon \end{pmatrix}, \quad \varepsilon \neq 0,$$

显然

$$A^+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$(A+E)^+ = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}.$$

易见$\lim\limits_{\varepsilon\to 0}(A+E)^{-1}$根本不存在. 但是,若

$$E=\begin{pmatrix}\varepsilon & 0\\ 0 & 0\end{pmatrix},$$

则当 $\varepsilon\to 0$时,

$$(A+E)^{+}=\begin{pmatrix}(1+\varepsilon)^{-1} & 0\\ 0 & 0\end{pmatrix}\to\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}=A^{+}.$$

注意这两种情况的差别. 对前一种情况,$r(A+E)>r(A)$. 而对后一种情况,$r(A+E)=r(A)$. 对这个特殊例子观察到的事实,对一般情况也是对的,这就是下面的定理.

定理4.7.1 (Stewart,1969)

$$\lim_{E\to 0}(A+E)^{+}=A^{+},$$

$\Longleftrightarrow E$ 接近于零时(即$\|E\|_2\to 0$时)$r(A+E)=r(A)$总成立.

为了证明本定理,我们需要如下三个引理.

引理4.7.1 设 $A\in C^{m\times n}$,$r(A+E)=r(A)$,

$$\|A^{+}\|_2\|E\|_2<1,$$

则

$$\|(A+E)^{-}\|_2\leqslant\frac{\|A^{+}\|_2}{1-\|A^{+}\|_2\|E\|_2}. \tag{4.7.2}$$

证明 因为 $r(A+E)=r(A)$,于是 $A+E$ 和 A 有相同个数的非零奇异值,设为 k 个,分别记为

$$\sigma_1\geqslant\cdots\geqslant\sigma_k,$$

$$\mu_1\geqslant\cdots\geqslant\mu_k.$$

于是利用定理2.10.3得

$$\mu_k\geqslant\sigma_k-\|E\|_2. \tag{4.7.3}$$

注意到

$$\sigma_k=\frac{1}{\|A^{+}\|_2},\quad \mu_k=\frac{1}{\|(A+E)^{+}\|_2},$$

代入(4.7.3)整理即得(4.7.2). 证毕.

引理4.7.2 设 $A\in C^{m\times n}$,记 $B=A+E$. 则

(a) $$B^{+}-A^{+}=-B^{+}EA^{+}+B^{+}(I_m-AA^{+})$$
$$-(I_n-B^{+}B)A^{+}.$$

(b) $$B^{+}-A^{+}=-B^{+}EA^{+}+B^{+}B^{*+}E^{*}(I_m-AA^{+})$$
$$-(I_n-B^{+}B)E^{*}A^{+*}A^{+}.$$

(c) $\|B^{+}-A^{+}\|_2\leqslant\frac{3}{2}(\|A^{+}\|_2^2+\|B^{+}\|_2^2)\|E\|_2.$

证明 (a) 的证明是容易的. 譬如将左端的 A^{+} 移至右端,然后验证右端是 B^{+}.

(b) 利用 $A^{*}AA^{+}=A^{*}$,则有

$$\begin{aligned}
B^{+}(I_m-AA^{+}) &= B^{+}BB^{+}(I_m-AA^{+})\\
&=B^{+}(BB^{+})^{*}(I_m-AA^{+})\\
&=B^{+}B^{*+}B^{*}(I_m-AA^{+})\\
&=B^{+}B^{*+}(A^{*}+E^{*})(I_m-AA^{+})\\
&=B^{+}B^{+*}E^{*}(I_m-AA^{+}).
\end{aligned}\qquad(4.7.4)$$

利用 $B^{+}BB^{*}=B^{*}$,同法可证

$$(I_n-B^{+}B)A^{+}=(I_n-B^{+}B)E^{*}A^{+*}A^{+}.\qquad(4.7.5)$$

将(4.7.4)和(4.7.5)代入(a),便得到(b).

(c) 将(b)的两端取范数,并利用

$$\|I_m-AA^{+}\|_2\leqslant 1,$$
$$\|I_n-B^{+}B\|_2\leqslant 1,$$

我们得到

$$\begin{aligned}
\|B^{+}-A^{+}\|_2 &\leqslant\|B^{+}\|_2\|E\|_2\|A^{+}\|_2+\|B^{+}\|_2^2\|E^{*}\|_2\\
&\quad+\|A^{+}\|_2^2\|E\|_2\\
&\leqslant(\|B^{+}\|_2\|A^{+}\|_2+\|B^{+}\|_2^2+\|A^{+}\|_2^2)\|E\|_2
\end{aligned}$$

$$\leqslant \frac{3}{2}(\|A^+\|_2^2 + \|B^+\|_2^2)\|E\|_2.$$

引理得证.

引理4.7.3 设 $A \in C^{m \times n}$，$r(A) = k < \min(m, n)$，若 $r(A+E) > r(A)$，则

$$\|(A+E)^+\|_2 \geqslant \frac{1}{\|E\|_2}.$$

证明 设 $r(A+E) = t > r(A) = k$. 则 A 的第 t 个奇异值 $\sigma_t = 0$，而 $A+E$ 的第 t 个奇异值 $\mu_t > 0$. 根据定理2.10.3，

$$\mu_t \leqslant \|E\|_2.$$

故有

$$\|(A+E)^+\|_2 = \frac{1}{\mu_t} \geqslant \frac{1}{\|E\|_2}.$$

引理证毕.

定理4.7.1的证明：

必要性 设 $A \in C^{m \times n}$，$r(A) = \min(m, n)$，则 $r(A+E) = r(A)$ 总成立. 记 $A = (a_1, \cdots, a_n)$，设 $r(A) = k < n$，且 A 的前 k 列线性无关. 记 $E = (\varepsilon_1, \cdots, \varepsilon_n)$. 容易证明，当 $\|\varepsilon_i\|$ 充分小时，$a_1 + \varepsilon_1, \cdots, a_k + \varepsilon_k$ 仍然线性无关，故 $r(A+E) \geqslant k = r(A)$.

若 $r(A+E) > r(A)$，依引理4.7.3，当 $\|E\|_2 \to 0$ 时，$\|(A+E)^+\|_2 \to \infty$，必要性得证.

充分性 设 $r(A+E) = r(A)$. 当 $\|E\|_2$ 充分小时，有

$$\|A^+\|_2 \|E\|_2 < 1.$$

由引理4.7.1知 $\|(A+E)^+\|_2$ 有上界. 再由引理4.7.2(c)，定理得证.

推论4.7.1 设 $A \in C^{m \times n}$ 为列满秩(即 $r(A) = n$)或行满秩(即 $r(A) = m$)，则

$$\lim_{E\to 0}(A+E)^{+}=A^{+},$$

即列满秩或行满秩矩阵的 Moore-Penrose 广义逆具有连续性.

推论4.7.2 设 $A\in C^{n\times n}$ 为可逆阵,则

$$\lim_{E\to 0}(A+E)^{-1}=A^{-1}.$$

即可逆阵具有连续性.

注 若 A 可逆,当 E 充分接近于零时,$A+E$ 也一定可逆,于是 $(A+E)^{-1}$ 总是存在的.

定理4.7.1表明,只有沿着能够保持 $A+E$ 与 A 具有相同秩的路径使 $E\to 0$,才能有

$$\lim_{E\to 0}(A+E)^{+}=A^{+}.$$

于是当 $r(A_{m\times n})=r<\min(m,n)$,$A^{+}$ 不具备连续性. 事实上,不论要求 $\|E\|$ 多小,总可以找到 E,使得

$$r(A+E)>r(A).$$

这个事实的证明如下:

记 $A=(a_1,\cdots,a_n)$,不妨设 $a_1,\cdots,a_r$ 线性无关. 任取向量 $b\overline{\in}\mathscr{M}(a_1,\cdots,a_r)$,定义

$$E=(0,\cdots,0,\varepsilon b,0,\cdots,0),\quad \varepsilon>0,$$

其中 εb 为 E 的第 $r+1$ 列. 随着 $\varepsilon\to 0$,显然 $\|E\|=\varepsilon\|b\|$ 可以任意小,但始终有

$$r(A+E)=r+1>r(A).$$

我们把上面讨论的结果写成如下定理.

定理4.7.2 设 $A\in C^{m\times n}$,$r(A)<\min(m,n)$,则 A^{+} 是不连续的,即

$$\lim_{E\to 0}(A+E)^{+}\neq A^{+}.$$

§4.8 最小二乘问题

假设 A 为 $m\times n$ 阵，b 为 $m\times 1$向量，若 $b\in\mathcal{M}(A)$，则线性方程组

$$Ax=b \tag{4.8.1}$$

是相容的. 我们已经证明了，它的全体解可表为

$$\{A^{(1)}b\},$$

这里 $A^{(1)}$为 A 的任意一个{1}-逆. 若 $b\overline{\in}\mathcal{M}(A)$，即(4.8.1)是不相容的，则对任意的 x，$Ax=b$ 都不成立，即残差向量 $r=Ax-b\neq 0$. 此时，我们退而求其次，寻找使残差向量最接近于零的 x，也就是，使 r 的范数

$$\|r\|=\|b-Ax\|$$

达到极小的 x. 我们称这样的 x 为(4.8.1)的最小二乘解. 又，在最小二乘解中，范数最小的解称为最小范数最小二乘解. 关于这一点，我们有如下定理：

定理4.8.1 线性方程组(4.8.1)的最小二乘解为

$$x=A^{+}b+(I_n-A^{+}A)z,\quad z\in C^n, \tag{4.8.2}$$

最小范数最小二乘解为

$$x_0=A^{+}b.$$

证明 把向量 b 作正交分解

$$b=b_1+b_2,$$

其中 $b_1=P_Ab$ 为 b 到 $\mathcal{M}(A)$的正交投影，$b_2=b-b_1$，于是

$$\|r\|^2=\|b-Ax\|^2=\|b_1-Ax\|^2+\|b_2\|^2,$$

可见，要使$\|r\|$达到极小，当且仅当 x 是方程组 $Ax=b_1$的解. 注意到 $b_1=AA^{+}b\in\mathcal{M}(A)$，故此方程组相容，将该方程组改写为

$$Ax = AA^+ b.$$

根据推论3.3.2知，它的通解为

$$\begin{aligned} x &= A^+ (AA^+ b) + (I - A^+ A)z \\ &= A^+ b + (I - A^+ A)z, \end{aligned}$$

其中 $z \in C^n$. (4.8.2)得证.

注意到 $x_0 = A^+ b$ 与 $(I - A^+ A)z$ 是正交的，所以

$$\|x\|^2 = \|x_0\|^2 + \|(I - A^+ A)z\|^2 \geqslant \|x_0\|^2$$

且等号成立 $\Longleftrightarrow (I - A^+ A)z = 0 \Longleftrightarrow x = x_0$. 定理证毕.

注 从证明过程，我们知道，不管方程组(4.8.1)是否相容，(4.8.2)总是它的某种意义上的通解. 若它相容，则(4.8.2)就是通常意义下的解. 若不相容，它就是最小二乘解.

推论4.8.1 若 $m \times n$ 矩阵 A 为列满秩，则 $x_0 = A^+ b$ 为(4.8.1)的唯一最小二乘解.

下面我们研究最小范数最小二乘解 $x_0 = A^+ b$ 的稳定性问题. 由于观测或测量误差或计算机的舍入误差，在(4.8.1)中矩阵 A 和常数向量 b 总是分别伴随有一定摄动 ΔA 和 Δb，此时，我们真正处理的方程组为

$$(A + \Delta A)(x + \Delta x) = b + \Delta b, \qquad (4.8.3)$$

它的最小范数解可表为

$$x_0 + \Delta x = (A + \Delta A)^+ (b + \Delta b), \qquad (4.8.4)$$

这里 $x_0 = A^+ b$. 当然我们关心 Δx 的大小，它度量了系数阵 A 和常数向量 b 的摄动对最小范数最小二乘解 x_0 的影响.

定理4.8.2 对方程组(4.8.3)，假设

(a) $r(A + \Delta A) = r(A)$;

(b) $d = \|A^+\|_2 \|\Delta A\|_2 < 1$,

则

$$\|\Delta x\| \leqslant \frac{\|A^+\|_2}{1-d}\Big(2\|\Delta A\|_2\|x_0\|$$

$$+ \|\Delta b\| + \frac{d}{1-d}\|b - Ax_0\|\Big).$$

证明 将 $x_0=A^+b$ 代入(4.8.4)，得

$$\Delta x = [(A + \Delta x)^+ - A^+]b + (A + \Delta x)^+ \Delta b.$$

记 $B=A+\Delta A, r_0=b-Ax_0$，利用引理4.7.2(b)，则有

$$\Delta x = - B^+ (\Delta A)x_0 + B^+ B^{+*} (\Delta A)^* r_0$$

$$- (I - B^+ B)(\Delta A)^* A^{-*} x_0$$

$$+ B^+ (\Delta b), \qquad (4.8.5)$$

由假设(b)，并应用(4.7.2)

$$\|B^+\|_2 \leqslant \frac{\|A^+\|_2}{1 - \|A^+\|_2\|\Delta A\|_2} \leqslant \frac{\|A^+\|_2}{1-d}. \qquad (4.8.6)$$

将(4.8.5)两端取欧氏范数，由范数相容性得

$$\|\Delta x\| \leqslant \|B^+\|_2\|\Delta A\|_2\|x_0\| + \|B^+\|_2^2\|\Delta A\|_2\|r_0\|$$

$$+ \|\Delta A\|_2\|A^+\|_2\|x_0\| + \|B^-\|_2\|\Delta b\|$$

$$= \|B^+\|_2(\|\Delta A\|_2\|x_0\| + \|\Delta b\|)$$

$$+ \|\Delta A\|_2\|A^+\|_2\|x_0\| + \|B^+\|_2^2\|\Delta A\|_2\|r_0\|.$$

利用(4.8.6)，得

$$\|\Delta x\| \leqslant \frac{\|A^+\|_2}{1-d}(\|\Delta A\|_2\|x_0\| + \|\Delta b\|)$$

$$+ \|\Delta A\|_2\|A^+\|_2\|x_0\| + \frac{\|A^+\|_2^2}{(1-d)^2}\|\Delta A\|_2\|r_0\|$$

$$= \frac{1}{1-d}\Big[2d\|x_0\| + \|A^+\|_2\|\Delta b\|$$

$$- d^2\|x_0\| + \frac{d\|A^+\|_2\|r_0\|}{1-d}\Big]$$

$$=\frac{\|A^+\|}{1-d}\Big[2\|\Delta A\|_2\|x_0\|+\|\Delta b\|$$

$$-\|A^+\|_2\|\Delta A\|_2^2\|x_0\|+\frac{d}{1-d}\|r_0\|\Big]$$

$$\leqslant\frac{\|A^+\|}{1-d}\Big[2\|\Delta A\|_2\|x_0\|+\|\Delta b\|+\frac{d}{1-d}\|r_0\|\Big].$$

定理证毕.

若(4.8.1)是相容的,则 $Ax_0=b$ 成立,于是我们有如下推论:

推论4.8.2 若(4.8.1)相容,则在定理假设下

$$\|\Delta x\|\leqslant\frac{\|A^+\|_2}{1-d}(2\|\Delta A\|_2\|x_0\|+\|\Delta b\|),$$

这里 d 的定义同定理4.8.2.

§4.9 加权 Moore-Penrose 广义逆

设在 C^m 和 C^n 中,分别定义了内积

$$(x,y)_M=y^*Mx,\quad\forall\, x,y\in C^m,\qquad(4.9.1)$$

$$(x,y)_N=y^*Nx,\quad\forall\, x,y\in C^n,\qquad(4.9.2)$$

其中 M、N 为 Hermite 正定阵,在§2.9中,对于 $m\times n$ 阵 A,我们定义了加权共轭转置阵

$$A^{\#}=N^{-1}AM.\qquad(4.9.3)$$

利用加权共轭转置阵,我们可以把 Moore-Penrose 广义逆加以推广.

定义4.9.1 设在 C^m 和 C^n 中,内积定义为(4.9.1)和(4.9.2),则对 $A\in C^{m\times n}$,若 $n\times m$ 阵 X 满足

$$AXA=A,$$

$$XAX=X,$$

$$(AX)^{\#}=AX,\quad (XA)^{\#}=XA, \tag{4.9.4}$$

则称 X 为 A 的加权 Moore-Penrose 广义逆，或最小 N 范数 M 最小二乘逆，记为 A_{MN}^{+}.

注 显然，当 M 和 N 皆为单位阵时，A_{MN}^{+}就变为通常的 A^{+}.

定理4.9.1 X 为 A_{MN}^{+}当且仅当 X 满足

$$\begin{aligned} AXA&=A,\\ XAX&=X,\\ (AX)^{*}M&=MAX,\\ (XA)^{*}N&=NXA. \end{aligned} \tag{4.9.5}$$

证明 利用(4.9.3)，即明所证.

定理4.9.2 A_{MN}^{+}是唯一的，且可表为

$$A_{MN}^{+}=N^{-\frac{1}{2}}(M^{\frac{1}{2}}AN^{-\frac{1}{2}})^{+}M^{\frac{1}{2}}.$$

证明 记 $B=M^{\frac{1}{2}}AN^{-\frac{1}{2}}$，$Y=N^{\frac{1}{2}}XM^{-\frac{1}{2}}$，则(4.9.5)的四个方程等价于

$$\begin{aligned} BYB&=B,\\ YBY&=Y,\\ (BY)^{*}&=BY,\\ (YB)^{*}&=YB. \end{aligned} \tag{4.9.6}$$

于是 $Y=B^{+}=(M^{\frac{1}{2}}AN^{-\frac{1}{2}})^{+}$，因而

$$X=A_{MN}^{+}=N^{-\frac{1}{2}}YM^{\frac{1}{2}}=N^{-\frac{1}{2}}(M^{\frac{1}{2}}AN^{-\frac{1}{2}})^{+}M^{\frac{1}{2}}.$$

定理证毕.

A_{MN}^{+}具有下列性质.

定理4.9.3 设 $A\in C^{m\times n}$，M、N 分别为 $m\times m$、$n\times n$ Hermite 正定阵，则

(a) $(A^+_{MN})^+_{NM}=A$.

(b) $(A^+_{MN})^*=(A^*)^+_{N^{-1}M^{-1}}$.

(c) $A^+_{MN}=(A^*MA)^+_{IN}A^*M$.

最后，我们给出 A^+_{MN} 在线性方程组求解中的应用.

在上一节，我们证明了，对于方程组 $Ax=b$，$x_0=A^+b$ 是最小范数最小二乘解，这里 A 为 $m\times n$ 阵，x 的范数 $\|x\|=(x^*x)^{\frac{1}{2}}$，即最小二乘是在范数 $\|b-Ax\|$ 达到最小意义下. 现在若取

$$\|x\|_N=(x^*Nx)^{\frac{1}{2}},\quad x\in C^n$$

并极小化

$$\|b-Ax\|_M=((b-Ax)^*M(b-Ax))^{1/2},$$

就导致了最小 N 范数 M 最小二乘解 $x_0=A^+_{MN}b$.

定理4.9.4 设 $A\in C^{m\times n}$，则 $x_0=A^+_{MN}b$ 是方程组 $Ax=b$ 最小 N 范数 M 最小二乘解.

证明 记 $\tilde{x}=N^{\frac{1}{2}}x$，$\tilde{b}=M^{\frac{1}{2}}b$，$\tilde{A}=M^{\frac{1}{2}}AN^{-\frac{1}{2}}$，则

$$\|x\|_N=\|\tilde{x}\|,$$

$$\|b-Ax\|_M=\|\tilde{b}-\tilde{A}\tilde{x}\|.$$

于是，方程组 $Ax=b$ 的最小 N 范数 M 最小二乘问题等价于方程组 $\tilde{A}\tilde{x}=\tilde{b}$ 的通常最小范数最小二乘问题. 由定理4.8.1知 $\tilde{x}=\tilde{A}^+\tilde{b}$ 为最小范数最小二乘解. 等价地，

$$\begin{aligned} x&=N^{-\frac{1}{2}}\tilde{A}^+\tilde{b}\\ &=N^{-\frac{1}{2}}(M^{\frac{1}{2}}AN^{-\frac{1}{2}})^+M^{\frac{1}{2}}b\\ &=A^+_{MN}b \end{aligned}$$

是原方程组的最小 N 范数 M 最小二乘解. 定理证毕.

第五章 其他$\{i,j,\cdots,l\}$-广义逆

前面两章，我们分别讨论了两种重要的广义逆，$\{1\}$-逆和$\{1,2,3,4\}$-逆. 本章将研究其余几种$\{i,j,\cdots,l\}$-逆.

§5.1 $\{1,2\}$-逆

设 A 为 $m\times n$ 矩阵，若 $n\times m$ 矩阵 X 满足 Penrose 方程(1.1.3)中的方程(1)和(2)，即

$$AXA=A, \tag{5.1.1}$$

$$XAX=X, \tag{5.1.2}$$

则称 X 为 A 的$\{1,2\}$-逆，记为 $A^{(1,2)}$，并用 $A\{1,2\}$表示它的全体. 早在50年代，统计学家 Rao 就对半正定方阵使用了这种广义逆，并称之为"伪逆"(Pseudo-inverse)，也有人称这种广义逆为自反广义逆，记之为 A_r^-. 这里 r 是reflexive g-inverse 的字头. 这是因为在(5.1.1)和(5.1.2)中，A 和 X 的地位完全对称，即若 $X\in A\{1,2\}$，则必有 $A\in X\{1,2\}$.

定理5.1.1 设 A 为 $m\times n$ 阵，且其相抵标准型为

$$A=p\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q, \tag{5.1.3}$$

其中 P、Q 分别为 $m\times m$ 和 $n\times n$ 可逆阵，$r=r(A)$，则

$$A^{(1,2)}=Q^{-1}\begin{pmatrix} I_r & B \\ C & CB \end{pmatrix}P^{-1},$$

这里 B 和 C 为适当阶数的任意阵.

证明 根据定义，$X\in A\{1,2\}\Longleftrightarrow X\in A\{1\}$、$XAX=X$，由

定理3.1.1，知

$$\Longleftrightarrow X = Q^{-1}\begin{pmatrix} I_r & B \\ C & D \end{pmatrix} P^{-1}, \quad XAX = X.$$

但 $XAX = X \Longleftrightarrow$

$$\begin{pmatrix} I_r & B \\ C & D \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_r & B \\ C & D \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} I_r & B \\ C & CB \end{pmatrix} = \begin{pmatrix} I_r & B \\ C & D \end{pmatrix}$$

$$\Longleftrightarrow D = CB,$$

定理证毕.

对任意矩阵 A，$A^{(1,2)}$也是 A 的一个 $A^{(1)}$，于是 $A^{(1,2)}$具有前面讨论过 $A^{(1)}$的所有性质.

定理5.1.2 $X \in A\{1,2\} \Longleftrightarrow$ 存在 $X_1, X_2 \in A\{1\}$，使得

$$X = X_1 A X_2.$$

证明 必要性 设 A 有相抵标准形(5.1.3)，根据定理5.1.1，若 $X \in A\{1,2\}$，则 X 可表为

$$X = Q^{-1}\begin{pmatrix} I_r & B \\ C & CB \end{pmatrix} P^{-1}$$

$$= Q^{-1}\begin{pmatrix} I_r & B_1 \\ C & D_1 \end{pmatrix} P^{-1} P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q Q^{-1}\begin{pmatrix} I_r & B \\ B_2 & D_2 \end{pmatrix} P^{-1}$$

$$= X_1 A X_2,$$

这里

$$X_1 = Q^{-1}\begin{pmatrix} I_r & B_1 \\ C & D_1 \end{pmatrix} P^{-1}, \quad X_2 = Q^{-1}\begin{pmatrix} I_r & B \\ B_2 & D_2 \end{pmatrix} P^{-1},$$

且 B_1, D_1和 B_2, D_2为任意阵. 容易验证 $X_i \in A\{1\}$，$i=1,2$.

充分性 $X = X_1 A X_2$，其中 $X_1, X_2 \in A\{1\}$，则

$$AXA = AX_1AX_2A = A,$$

$$XAX = X_1AX_2AX_1AX_2 = X_1AX_2 = X.$$

定理证毕.

定理5.1.3 对给定的矩阵 A 和 $X \in A\{1\}$,则

$$X \in A\{1,2\} \Longleftrightarrow r(X) = r(A).$$

证明 必要性 $r(A) = r(AXA) \leqslant r(X) = r(XAX) \leqslant r(A)$.

充分性 $\because X \in A\{1\}, \therefore r(XA) = r(A) = r(X)$. 但 $\mathscr{M}(XA) \subset \mathscr{M}(X)$,于是 $\mathscr{M}(XA) = \mathscr{M}(X)$,存在 Y,使得 $X = XAY$,用 A 左乘,得

$$AX = AXAY = AY.$$

再用 X 左乘,

$$XAX = XAY = X.$$

定理证毕.

定理5.1.4 设 P,Q 为可逆阵,则

$$(PAQ)^{(1,2)} = Q^{-1}A^{(1,2)}P^{-1}.$$

证明是容易的.

为了讨论具有指定列空间和零空间的{1,2}-逆的构造,我们先证明一个引理.

引理5.1.1 对任意矩阵 A,至多有一个 $A^{(2)}$满足下面的矩阵方程:

$$AX = B, \quad XA = C.$$

证明 问题等价于证明矩阵方程

$$AX = B, XA = C, XAX = X \tag{5.1.4}$$

至多有一个公共解. 假定(5.1.4)有解,且 X_1和 X_2为两个解,记

$$W = X_1 - X_2,$$

则

$$AW = 0, WA = 0, WB = W, CW = W.$$

于是

$$W^*W = W^*C^*WB = W^*A^*X_i^*WAX_i = 0, i = 1,2.$$

这表明 $W=0$，故有 $X_1=X_2$. 引理得证.

定理5.1.5 设 $A\in C^{m\times n}$，$\mathscr{R}(A)=L$，$\mathscr{N}(A)=M$，$L\oplus S=C^m$，$T\oplus M=C^n$，则 $A\{1,2\}$中以 T 为列空间，S 为零空间的矩阵是唯一的，且为

$$G = P_{T,M}A^- P_{L,S}, \tag{5.1.5}$$

这里 $P_{T,M}$为循 M 向 T 的投影阵.

证明 首先，由引理3.6.1知，(5.1.5)定义的 G 满足 $AGA=A$，于是 $G\in A\{1\}$，$r(G)\geqslant r(A)$. 另一方面

$$\begin{aligned} r(G) &= r(P_{T,M}A^- P_{L,S}) \leqslant r(P_{L,S}) \\ &= r(AA^-) = r(A), \end{aligned}$$

故有 $r(G)=r(A)$. 注意到

$$\mathscr{R}(GA)\subset\mathscr{R}(G),$$

结合 $r(GA)=r(A)=r(G)$，我们有 $\mathscr{R}(GA)=\mathscr{R}(G)$，因而存在 U，使得 $G=GAU$，于是

$$AG = AGAU = AU,$$

$$GAG = GAU = G, \tag{5.1.6}$$

这表明 $G\in A\{2\}$，即证明了 $G\in A\{1,2\}$.

利用引理3.6.1，不难推出

$$P_{L,S}A = A,$$

$$AP_{T,M} = A.$$

于是

$$AG = AA^- P_{L,S} = P_{L,S}, \tag{5.1.7}$$

$$GA = P_{T,M}A^- A = P_{T,M}. \tag{5.1.8}$$

由引理5.1.1知，满足(5.1.6)，(5.1.7)和(5.1.8)的 G 最多只有一个，这就是(5.1.5)定义的 G.

最后，因为 $P_{L,S}=AA^{-}$，$P_{T,M}=A^{-}A$，于是

$$\mathscr{M}(G)=\mathscr{M}(GA)=\mathscr{M}(P_{T,M})=T,$$

$$\mathscr{N}(G)=\mathscr{N}(AG)=\mathscr{N}(P_{L,S})=S.$$

定理证毕.

注 定理5.1.5表明，(5.1.5)定义的 G 是唯一一个具有指定列空间 T 和零空间 S 的 $A\{1,2\}$，以后，把它简记为 $A_{T,S}^{(1,2)}$.

下面的定理提供了计算 $A_{T,S}^{(1,2)}$ 的一个公式，称为 Urquhart 公式.

定理5.1.6 设 A 为 $m\times n$ 阵，P 和 Q 分别为 $n\times p$，$q\times m$ 矩阵，记

$$X=P(QAP)^{-}Q,$$

则

$$X=A_{\mathscr{M}(P),\mathscr{N}(Q)}^{(1,2)}\Longleftrightarrow r(QAP)=r(P)=r(Q)=r(A).$$

证明 根据定理3.6.2，我们有

$$X\in A\{1\}\Longleftrightarrow r(QAP)=r(A).$$

又根据定理5.5.3，

$$X\in A\{2\},\mathscr{M}(X)=\mathscr{M}(P)\Longleftrightarrow r(QAP)=r(P),$$

$$X\in A\{2\},\mathscr{N}(X)=\mathscr{N}(Q)\Longleftrightarrow r(QAP)=r(Q).$$

于是结论得证.

下面的定理刻画了 A^{+} 与{1,2}-逆的关系.

定理5.1.7 设 A 为 $m\times n$ 矩阵，则

$$A^{+}=A_{\mathscr{M}(A^{*}),\mathscr{N}(A^{*})}^{(1,2)}.$$

证明 显然，$A^{+}\in A\{1,2\}$，再由定理4.2.4和定理4.2.5

知：

$$\mathcal{N}(A^{+}) = \mathcal{N}(A^{*}),$$

$$\mathcal{M}(A^{+}) = \mathcal{M}(A^{*}),$$

结合定理5.1.5，命题得证.

这个定理表明，在广义逆集 $A\{1,2\}$ 中，存在唯一的 $A^{(1,2)}$，满足 $\mathcal{M}(A^{(1,2)})=\mathcal{M}(A^{*})$，$\mathcal{N}(A)=\mathcal{N}(A^{*})$，这个 $A^{(1,2)}$ 就是 A^{+}.

§5.2 {1,3}-逆

设 A 为 $m\times n$ 矩阵，我们称满足(1.1.3)中方程(1)和(3)，即

$$AXA = A,$$
$$(AX)^{*} = AX$$

的 $n\times m$ 矩阵 X 为 A 的{1,3}-逆，记为 $A^{(1,3)}$，并用 $A\{1,3\}$ 表示它的全体. 这种广义逆又称为最小二乘广义逆，因此有的文献把它记为 A_l^{-}，这里 l 是"最小二乘"(least square)的英文字头.

下面的引理给出了 $A\{1,3\}$ 一个表征.

引理5.2.1 设 $G_0\in A\{1,3\}$，则集合 $A\{1,3\}$ 为矩阵方程

$$AX = AG_0 \tag{5.2.1}$$

的全体解.

证明 先证明(5.2.1)的任一解都是一个 $A^{(1,3)}$. 设 X_0 满足(5.2.1)，则

$$AX_0A = AG_0A = A,$$

且因 AG_0 是 Hermite 阵，故 $(AX_0)^{*}=(AG_0)^{*}=AG_0=AX_0$. 这

就证明了 $X_0\in A\{1,3\}$.

反之，对任一 $X\in A\{1,3\}$，则

$$AG_0 = AXAG_0 = (AX)^*(AG_0)^* = X^*A^*G_0^*A^*$$
$$= X^*(AG_0A)^* = X^*A^* = (AX)^* = AX,$$

引理证毕.

定理5.2.1 (a) $AA^{(1,3)}=AA^+=P_A$.

(b) $AA^{(1,3)}$为 Hermite 幂等阵.

证明 (a) 在(5.2.1)中取 $G_0=A^+$，结合定理4.2.4即明所证.

(b) 是(a)的直接推论.

利用引理5.2.1我们立即得到 $A\{1,3\}$的如下表征.

定理5.2.2 设 A 为 $m\times n$ 矩阵，$G_0\in A\{1,3\}$，则

$$A\{1,3\} = \{G_0 + (I_n - G_0A)Z, Z\text{ 为任一 } n\times m \text{ 阵}\}.$$

证明 在定理3.3.1中，取 $B=I, H=AG_0, Y=Z+G_0$，再利用引理5.2.1即得所要结论.

$A^{(1,3)}$之所以称为 A 的最小二乘广义逆，是源于如下事实.

定理5.2.3 $x=Gb$ 为不相容线性方程组 $Ax=b$ 的最小二乘解的充要条件为 $G\in A\{1,3\}$.

证明 由定理4.8.1的证明过程知，x 为 $Ax=b$ 的最小二乘解，当且仅当

$$Ax = AA^+b = P_Ab, \qquad (5.2.2)$$

因此，若 $x_0=A^{(1,3)}b$，利用定理5.2.1得

$$Ax_0 = AA^{(1,3)}b = P_Ab.$$

反过来，若 $x=Gb$ 是最小二乘解，由(5.2.2)知，对一切 $b\in C^m$，

$$AGb = P_Ab,$$

故 $AG=P_A=AA^{(1,3)}$，由引理5.2.1知，$G\in A\{1,3\}$，证毕.

综合上述两个定理，我们有

推论5.2.1 方程组 $Ax=b$ 的最小二乘解可表为

$$x = A^{(1,3)}b + (I - A^{(1,3)}A)y. \tag{5.2.3}$$

注 读者试比较(5.2.3)与(4.8.2). 利用引理5.2.1知，两者是等价的.

定理5.2.4 设 A 为 $m\times n$ 矩阵，则 $X\in A\{1,3\}\Longleftrightarrow X$ 为矩阵方程 $AX=I_m$ 的最小二乘解，即 X 极小化 $\|AX-I\|$.

(注，此处 $\|A\|$ 表示矩阵 A 的欧氏范数，参见 §2.9.)

证明 记 $\mathrm{Vec}(A)$ 为将矩阵 A 的列向量依次排成的列向量，$A\otimes B$ 表示矩阵 A 与 B 的 Kronecker 乘积(参阅王松桂，1987，p. 37). 对 $AX=I_m$ 两边做向量化运算，得

$$(I\otimes A)\mathrm{Vec}(X) = \mathrm{Vec}(I_m).$$

从上一推论知，此方程组最小二乘解为

$$\mathrm{Vec}(X) = (I\otimes A)^{(1,3)}\mathrm{Vec}(I_m) + (I - (I\otimes A)^{(1,3)}(I\otimes A))z,$$

$$z\in C^{mn},$$

应用定义，容易验证

$$I\otimes A^{(1,3)} \in (I\otimes A)\{1,3\},$$

于是

$$\mathrm{Vec}(X) = (I\otimes A^{(1,3)})\mathrm{Vec}(I_m) + (I - (I\otimes A^{(1,3)}A))z,$$

即

$$X = A^{(1,3)} + (I - A^{(1,3)}A)Z,$$

Z 为 $n\times m$ 矩阵，由定理5.2.2，明所欲证.

上面诸结果表明，$\{1,3\}$-逆与最小二乘解有着密切的联系. 因此，凡在用到最小二乘法的情况，$\{1,3\}$-逆都起着重要

作用.

最后,我们把{1,3}-逆推广到加权{1,3}-逆.

若在 C^m 中,定义内积

$$(x,y)_M = y^* Mx, \quad \forall\, x,y \in C^m,$$

其中 M 为 Hermite 正定阵. 我们定义 A 的加权{1,3}-逆为满足

$$\begin{aligned} AXA &= A, \\ (AX)^\# &= AX \end{aligned} \tag{5.2.4}$$

的矩阵 X,记为 $A^{(1,3)}_{(M)}$,它的全体记为 $A_{(M)}\{1,3\}$,对应于记号 A_l^-,加权 $A^{(1,3)}_{(M)}$ 也记为 $A^-_{l(M)}$. 容易证明(5.2.4)等价于

$$\begin{aligned} AXA &= A, \\ (AX)^* M &= MAX. \end{aligned} \tag{5.2.5}$$

易见,当 $M=I$ 时,$A^{(1,3)}_{(M)}$ 就化为 $A^{(1,3)}$.

对方程组 $Ax=b$,在极小化范数

$$\|Ax-b\|_M = ((b-Ax)^* M(b-Ax))^{1/2}$$

意义下的最小二乘解,称为 M 最小二乘解(或加权最小二乘解).

定理5.2.5 $x=Gb$ 为不相容线性方程组 $Ax=b$ 的 M 最小二乘解 $\Longleftrightarrow G\in A_{(M)}\{1,3\}$.

证明 记 $\tilde{x}=M^{\frac{1}{2}}x$, $\tilde{b}=M^{\frac{1}{2}}b$, $\tilde{A}=M^{\frac{1}{2}}AM^{-\frac{1}{2}}$,则

$$\|Ax-b\|_M = \|\tilde{A}\tilde{x}-\tilde{b}\|.$$

于是,方程组 $Ax=b$ 的 M 最小二乘解问题就等价于方程组 $\tilde{A}\tilde{x}=\tilde{b}$ 的普通最小二乘解问题. 依定理5.2.3, $\tilde{x}_0=\tilde{G}\tilde{b}$ 为最小二乘解 $\Longleftrightarrow$

$$\tilde{G}\in\tilde{A}\{1,3\},$$

即 $\tilde{G}$ 满足

$$\begin{cases}\tilde{A}\tilde{G}\tilde{A}=\tilde{A},\\(\tilde{A}\tilde{G})^*=\tilde{A}\tilde{G}.\end{cases}$$

记 $G=M^{-\frac{1}{2}}\tilde{G}M^{\frac{1}{2}}$，上面两式等价于

$$AGA=A,$$
$$(AG)^*M=MAG,$$

即 $G\in A_{(M)}\{1,3\}$，于是，从 $\tilde{x}_0=\tilde{G}\tilde{b}$，得

$$x_0=M^{-\frac{1}{2}}\tilde{G}\tilde{b}=M^{-\frac{1}{2}}\tilde{G}M^{\frac{1}{2}}b=Gb,$$

定理证毕.

§5.3 {1,4}-逆

设 A 为 $m\times n$ 阵，若 $n\times m$ 阵 X 满足(1.1.3)中的方程(1)和(4)，即

$$AXA=A,$$
$$(XA)^*=XA,$$

则称 X 为 A 的{1,4}-逆，记为 $A^{(1,4)}$. 记 $A\{1,4\}$ 为全体的 $A^{(1,4)}$ 组成的集合. 因为这种广义逆又称为最小范数广义逆，因此，有的文献把它记为 A_m^-(参阅 Rao 等，1971，p. 45).

引理5.3.1 设 $G_0\in A\{1,4\}$，则集合 $A\{1,4\}$ 为矩阵方程

$$XA=G_0A \tag{5.3.1}$$

的全体解.

证明 先证(5.3.1)的任一解都是一个 $A^{(1,4)}$. 设 X_0 为(5.3.1)的一个解，则

$$AX_0A=AG_0A=A.$$

又因 G_0A 为 Hermite 阵，故 $(X_0A)^*=(G_0A)^*=G_0A=X_0A$. 这就证明了 $X_0\in A\{1,4\}$.

反过来，$\forall X\in A\{1,4\}$，则

$$\begin{aligned} G_0A &= G_0AXA = (G_0A)^*(XA)^* \\ &= A^*G_0^*A^*X^* = (AG_0A)^*X^* \\ &= A^*X^* = (XA)^* = XA. \end{aligned}$$

引理证毕.

定理5.3.1 (a) $A^{(1,4)}A=A^+A=P_{A^*}$.

(b) $A^{(1,4)}A$ 为 Hermite 幂等阵.

证明 (a) 在(5.3.1)中取 $G_0=A^+$，结合定理4.2.5，即得欲证.

(b) 是(a)的直接推论.

定理5.3.2 设 A 为 $m\times n$ 阵，$G_0\in A\{1,4\}$，则

$$A\{1,4\}=\{G_0+Z(I_m-AG_0),\ Z\ 为任一\ m\times n\ 阵\}.$$

证明 在定理3.3.1中，取 $A=I,H=G_0A,Y=Z+G_0$，利用引理5.3.1，命题得证.

$A^{(1,4)}$又称为 A 的最小范数广义逆，此名源于如下事实.

定理5.3.3 设 A 为 $m\times n$ 阵，且 $Ax=b$ 为相容线性方程组. $G\in A\{1\}$，则对一切 $b\in\mathscr{M}(A)$，$x_0=Gb$ 为最小范数解$\Longleftrightarrow$ $G\in A\{1,4\}$.

证明 $\Leftarrow$ 设 $x_0=A^{(1,4)}b$，对 $b\in\mathscr{M}(A)$，存在 t，使得 $b=At$，于是 $x_0=A^{(1,4)}At\in\mathscr{M}(A^*)$，由 $Ax=b$ 的通解表示

$$\begin{aligned} x &= A^{(1,4)}b+(I-GA)u \\ &= x_0+(I-GA)u,\quad u\in C^n. \end{aligned}$$

注意到

$$(A^{(1,4)}A)^*(I-GA)=0,$$

故

$$\|x\|=\|x_0\|+\|(I-GA)u\|\geqslant\|x_0\|,$$

等号成立$\Longleftrightarrow x=x_0$.

于是我们证明了，对给定的 b，$x_0\in\mathscr{R}(A^*)$的最小范数解是唯一的，且为 $x_0=A^{(1,4)}b$.

$\Rightarrow$ 若对一切 $b\in\mathscr{R}(A)$，$x=Gb$ 为 $Ax=b$ 的最小范数解. 记 $A=(a_1,\cdots,a_n)$，依次取 $b=a_i$，$i=1,\cdots,n$，由刚才证明过的唯一性，知

$$Ga_i=A^{(1,4)}a_i,\quad i=1,2,\cdots,n,$$

即

$$GA=A^{(1,4)}A.$$

由引理5.3.1知，$G\in A\{1,4\}$. 定理证毕.

下面的定理给出了 $A\{1,4\}$的另一种刻画.

定理5.3.4 设 $A\in C^{m\times n}$，则 $X\in A\{1,4\}\Longleftrightarrow X$ 为矩阵方程

$$XA=I_n \tag{5.3.2}$$

的最小二乘解.

证明 对(5.3.2)两边取共轭，得

$$A^*X^*=I_n. \tag{5.3.3}$$

利用定理5.2.4，这个矩阵方程的最小二乘解为

$$X^*\in A^*\{1,3\},$$

这等价于 $X\in A\{1,4\}$，定理证毕.

下面的定理是利用 $A^{(1,3)}$和 $A^{(1,4)}$给出 A^+的一种表征.

定理5.3.5 $A^+=A^{(1,4)}AA^{(1,3)}$.

证明 利用定理5.2.1和定理5.3.1，我们有

$$A^{(1,4)}AA^{(1,3)}=A^+AA^{(1,3)}=A^+AA^+=A^+.$$

定理证毕.

最后，我们把{1,4}-逆推广到加权{1,4}-逆.

若在 C^n 中，定义内积

$$(x,y)_N = y^* N x, \quad \forall\, x, y \in C^n,$$

其中 N 为 Hermite 正定阵. 我们定义 A 的加权 $\{1,4\}$-逆为满足

$$AXA = A,$$
$$(XA)^{\#} = XA$$

的矩阵 X, 记为 $A_{(N)}^{(1,4)}$, 它们的全体记为 $A_{(N)}\{1,4\}$. 对应于记号 A_m^-, 加权 $A_{(N)}^{(1,4)}$ 也记为 $A_{m(N)}^-$. 不难证明, 上面两个矩阵方程等价于

$$AXA = A,$$
$$(XA)^* N = NXA.$$

易见, 当 $N=I$ 时, $A_{(N)}^{(1,4)}$ 就化为 $A^{(1,4)}$.

对于相容线性方程组 $Ax=b$, 在所有解 $x=Gb$ 中, 其中 $G\in A\{1\}$; 我们考虑范数 $\|x\|_N$ 最小者, 称为最小 N 范数解.

定理5.3.6 设 A 为 $m\times n$ 阵, 对 $G\in A\{1\}$, 相容线性方程组 $Ax=b$, 对一切 $b\in \mathscr{M}(A)$, $x_0=Gb$ 为最小 N 范数解 $\Longleftrightarrow$ $G\in A_{(N)}\{1,4\}$.

证明 命 $\tilde{x}=N^{\frac{1}{2}}x$, $\tilde{A}=AN^{-\frac{1}{2}}$, 则 $Ax=b$ 的最小 N 范数解问题等价于 $\tilde{A}\tilde{x}=b$ 的最小范数解问题. 根据定理5.3.3, $\tilde{G}\in \tilde{A}\{1\}$, 对一切 $b\in \mathscr{M}(\tilde{A})=\mathscr{M}(A)$, $\tilde{x}_0=\tilde{G}b$ 为 $\tilde{A}\tilde{x}=b$ 的最小范数解 $\Longleftrightarrow \tilde{G}\in \tilde{A}\{1,4\} \Longleftrightarrow \tilde{G}$ 满足

$$\begin{cases} \tilde{A}\tilde{G}\tilde{A} = \tilde{A}, \\ (\tilde{G}A)^* = \tilde{G}\tilde{A}. \end{cases}$$

令 $G=N^{-\frac{1}{2}}\tilde{G}$, 上式等价于

$$\begin{cases} AGA = A, \\ (GA)^* N = NGA. \end{cases}$$

于是 $G \in \tilde{A}_{(N)}\{1,4\}$，故

$$x_0 = N^{-\frac{1}{2}}\tilde{x}_0 = N^{-\frac{1}{2}}\tilde{G}b = Gb,$$

定理证毕.

§ 5.4 {1,2,3}与{1,2,4}-逆

设 A 为 $m\times n$ 阵，若 X 满足(1.1.3)中的方程(1)、(2)和(3)，即

$$AXA = A, \tag{5.4.1}$$

$$XAX = X, \tag{5.4.2}$$

$$(AX)^* = AX, \tag{5.4.3}$$

则称 X 为 A 的{1,2,3}-逆，记为 $A^{(1,2,3)}$，这样的矩阵的全体记为 $A\{1,2,3\}$. 若把上面的第三个方程代之以 $(XA)^* = XA$，相应的 X，称为 A 的{1,2,4}-逆，我们有相应的记号 $A^{(1,2,4)}$ 和 $A\{1,2,4\}$.

定理5.4.1 对任意 $m\times n$ 矩阵 A，

(a) $n\times m$ 阵 X 为 A 的一个{1,2,3}-逆 $\Longleftrightarrow X = (A^*A)^- A^*$.

(b) $n\times m$ 阵 X 为 A 的一个{1,2,4}-逆 $\Longleftrightarrow X = A^*(AA^*)^-$.

证明 (a) $\Leftarrow$ 设 $X = (A^*A)^- A^*$，则

$$AX = A(A^*A)^- A^* = AA^+.$$

显然有 $AXA = A$，从上式知 AX 为向 $\mathscr{M}(A)$ 的正交投影阵，即 $AX = P_A$，于是它是 Hermite 阵. 从而(5.4.3)得证. 另外，

$$A^*AA^+ = A^*P_A = (P_A A)^* = A^*,$$

于是 $XAX = XAA^+ = (A^*A)^- A^*AA^+ = (A^*A)^- A^* = X.$

(5.4.2)得证.

$\Rightarrow$ 若 $X\in A\{1,2,3\}$，则

$$X = XAX = X(AX)^* = XX^*A^*.$$

只需证明，$XX^*\in(A^*A)\{1\}$.

事实上，

$$A^*A(XX^*)A^*A = A^*P_AA = A^*A.$$

(b) 证明与(a)相类似，留给读者作练习.

推论5.4.1 (a) $AA^{(1,2,3)}=AA^+=P_A$.

(b) $A^{(1,2,4)}A=A^+A=P_{A^*}$.

推论5.4.2 (a) $A^*AA^{(1,2,3)}=A^*$.

(b) $A^{(1,2,4)}AA^*=A^*$.

下面的定理是用{1,2,3}-逆和{1,2,4}-逆来刻画乘积阵的{1}-逆.

定理5.4.2 设 P_1 和 P_2 分别是 $l\times m$ 和 $n\times q$ 矩阵，且 $r(P_1)=m$，$r(P_2)=n$，A 为 $m\times n$ 阵，则对任一个 A^-，有

(a) $(P_1AP_2)^-=P_2^{(1,2,4)}A^-P_1^{(1,2,3)}$.

(b) $A^-=P_2(P_1AP_2)^-P_1$.

证明 (a)因为 P_1 是列满秩，所以 $P_1^{(1,2,3)}$ 是 P_1 的左逆阵，而 P_2 是行满秩，因此，$P_2^{(1,2,4)}$ 是 P_2 的右逆阵，则由 $AA^-A=A$ 得到

$$P_1AP_2(P_2^{(1,2,4)}A^-P_1^{(1,2,3)})P_1AP_2 = P_1AP_2.$$

(a)得证.

(b) 将表达式 $(P_1AP_2)^-=P_2^{(1,2,4)}A^-P_1^{(1,2,3)}$ 两边左乘 P_2，右乘 P_1，利用 $P_1^{(1,2,3)}P_1=I$，$P_2P_2^{(1,2,4)}=I$，命题得证.

类似地，可以证明

推论5.4.3 在定理5.4.2假设下

(a) $(P_1AP_2)^{(1,2)}=P_2^{(1,2,4)}A^{(1,2)}P_1^{(1,2,3)}$.

(b) $A^{(1,2)}=P_2(P_1AP_2)^{(1,2)}P_1$.

定理5.4.3 设 P_2和 A 同定理5.4.2，$P_1^*P_1=I_m$. 则

(a) $(P_1AP_2)^{(1,2,3)}=P_2^{(1,2,4)}A^{(1,2,3)}P_1^*$.

(b) $A^{(1,2,3)}=P_2(P_1AP_1)^{(1,2,3)}P_1$.

证明 (a) 由定理5.4.1，

$$\begin{aligned}(P_1AP_2)^{(1,2,3)} &=((P_1AP_2)^*(P_1AP_2))^-(P_1AP_2)^* \\ &=(P_2^*A^*AP_2)^-P_2^*A^*P_1^*. \qquad (5.4.4)\end{aligned}$$

因为 P_2有右逆阵 $P_2^{(1,2,4)}$，P_2^* 有左逆阵 $P_2^{(1,2,3)^*}$，应用定理5.4.2，

$$(P_2^*A^*AP_2)^-=P_2^{(1,2,4)}(A^*A)^-P_2^{(1,2,3)^*}.$$

代入(5.4.4)得

$$\begin{aligned}&P_2^{(1,2,4)}(A^*A)^-P_2^{(1,2,3)^*}P_2^*A^*P_1^* \\ &=P_2^{(1,2,4)}(A^*A)^-A^*P_1^* \\ &=P_2^{(1,2,4)}A^{(1,2,3)}P_1^*\in(P_1AP_2)\{1,2,3\}.\end{aligned}$$

(a)得证.

(b) 证明是容易的，留作练习.

上述两个定理，是求分块矩阵的广义逆的有力工具.

定理5.4.4 下列三个命题是等价的：

(a) $A^*B=0$;

(b) $A^{(1,2,3)}B=0$;

(c) $B^{(1,2,3)}A=0$.

证明 由推论5.4.2得

$$A^*B=A^*AA^{(1,2,3)}B,$$

于是我们证明了(b)⇒(a). 另一方面，由定理5.4.1，

$$A^{(1,2,3)}B=(A^*A)^-A^*B.$$

于是，(a)⇒(b)，这就证明了(a)与(b)的等价性. 从已证结果

交换字母 A,B 可得(a)与(c)的等价性.定理证毕.

§5.5 {2}-逆

设 A 为 $m\times n$ 阵,若 $n\times m$ 阵 X 满足(1.1.3)中的第二个方程,即

$$XAX = X, \tag{5.5.1}$$

则称 X 为 A 的一个{2}-逆,记为 $A^{(2)}$,这样的矩阵的全体记为 $A\{2\}$.下面我们将应用满秩分解来构造 $A^{(2)}$.

在§5.1,我们已经给出了 $A^{(1,2)}$ 的构造,于是对给定的 A,假设有了 $X_0\in A\{1,2\}$,给定$1\leqslant t\leqslant r$,$r(A)=r$,我们要构造一个秩为 t 的 $A^{(2)}$.对 X_0作满秩分解

$$X_0 = YZ,$$

其中 Y 和 Z 分别为 $n\times r$ 和 $r\times m$ 阵,$r(Y)=r(Z)=r=r(A)$.X_0也是一个 $A^{(2)}$,故满足(5.5.1),即

$$YZAYZ = YZ.$$

用 Y^- 和 Z^-分别左乘和右乘上式两边,应用定理3.2.2,得

$$ZAY = I_r. \tag{5.5.2}$$

分别对 Y,Z 作分块

$$Y = (Y_1 \vdots Y_2),\quad Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

其中 Y_1和 Y_2分别为 $n\times t$ 和 $n\times(r-t)$矩阵,而 Z_1和 Z_2分别为 $t\times m$ 和$(r-t)\times m$ 矩阵,由(5.5.2)得

$$Z_1AY_1 = I_t. \tag{5.5.3}$$

记 $X_1=Y_1Z_1$,在(5.5.3)两边,左乘 Y_1,右乘 Z_1,得

$$X_1AX_1 = X_1, \tag{5.5.4}$$

即 X_1为一个 $A^{(2)}$,且 $r(X_1)=r(Y_1Z_1)=r(Y_1)=r(Z_1)=t$.这就

是说，X_1是一个秩为 t 的 $A^{(2)}$.

我们把上面证明的结果，写成如下定理.

定理5.5.1 设 A 为 $m\times n$ 阵，$r(A)=r$. 给定正整数 t，$1\leqslant t\leqslant r$. 设 X_0为一个 $A^{(1,2)}$，对 X_0作满秩分解 $X_0=YZ$，设 Y_1和 Z_1分别为 Y 的前 t 列和 Z 的前 t 行，则 $X_1=Y_1Z_1\in A\{2\}$，且 $r(X_1)=t$.

在具有给定秩的 $A^{(2)}$构造中，方程(5.5.3)起着特别重要的作用. 事实上，仅条件(5.5.3)就足以刻画具有给定秩的 $A^{(2)}$，这就是下面的定理.

定理5.5.2 设 A 为 $m\times n$ 阵，$r(A)=r$，对正整数 t，$1\leqslant t\leqslant r$，用 $A\{2\}_t$ 表示 $A\{2\}$中秩为 t 的那些矩阵组成的集合，则

$$A\{2\}_t = \{YZ: Y \text{ 为 } n\times t \text{ 阵}, Z \text{ 为 } t\times m \text{ 阵}, ZAY = I_t\}. \tag{5.5.5}$$

证明 记

$$S = \{YZ: Y \text{ 为 } n\times t \text{ 阵}, Z \text{ 为 } t\times m \text{ 阵}, ZAY = I_t\}.$$

先证

$$S \subset A\{2\}_t. \tag{5.5.6}$$

设 $X\in S$，则 X 可分解为

$$X = YZ, \tag{5.5.7}$$

其中 Y 为 $n\times t$ 阵，Z 为 $t\times m$ 阵，且

$$ZAY = I_t. \tag{5.5.8}$$

从上式易知，$r(Y)=r(Z)=t$，从(5.5.7)知，$r(X)=t$，从而

$$XAX = YZAYZ = X.$$

这就证明了 $X\in A\{2\}_t$，于是(5.5.6)得证.

再证 $A\{2\}_t\subset S$. 设 $X\in A\{2\}_t$，且有满秩分解

$$X = YZ,$$

其中 Y 为 $n\times t$ 阵，Z 为 $t\times m$ 阵，$r(Y)=r(Z)=t$，则

$$YZAYZ = YZ. \tag{5.5.9}$$

因为 Y 为列满秩，Z 为行满秩，由定理3.2.2，对任意 Y^- 和 Z^- 有

$$Y^- Y = ZZ^- = I_t.$$

用 Y^- 左乘，Z^- 右乘(5.5.9)两边，便有

$$ZAY = I_t.$$

这就证明了 $X\in S$，于是 $A\{2\}_t\subset S$，证毕.

下面的定理平行于定理3.6.2，它给出了具有给定列空间和零空间的{2}-逆构造方法.

定理5.5.3 设 A 为 $m\times n$ 阵，P 和 Q 分别为 $n\times p$，$q\times m$ 阵，记

$$X = P(QAP)^- Q, \tag{5.5.10}$$

其中，$(QAP)^-$ 为任意一个 $(QAP)^{(1)}$. 则

(a) $X\in A\{2\}$，$\mathscr{M}(X)=\mathscr{M}(P)\Leftrightarrow r(QAP)=r(P)$.

(b) $X\in A\{2\}$，$\mathscr{N}(X)=\mathscr{N}(Q)\Leftrightarrow r(QAP)=r(Q)$.

证明 (a) 充分性 假设 $r(QAP)=r(P)$，依引理3.6.2，有

$$XAP = P(QAP)^- QAP = P,$$

由此立即得

$$XAX = P(QAP)^- QAP(QAP)^- Q = P(QAP)^- Q = X.$$

于是 $X\in A\{2\}$ 且 $r(X)=r(P)$，据此易证 $\mathscr{M}(X)=\mathscr{M}(P)$.

必要性 设 $X\in A\{2\}$，则

$$X = XAX = P(QAP)^- QAP(QAP)^- Q,$$

故

$$r(X)\leqslant r(QAP)\leqslant r(P).$$

根据 $\mathscr{M}(X)=\mathscr{M}(P)$，我们有 $r(X)=r(P)$，结合上式便证明了所要结论.

(b) 证明类似于(a)，留给读者作练习.

推论5.5.1 设 $A\in C^{m\times n}$，$P\in C^{n\times p}$，$Q\in C^{q\times m}$，记

$$X = P(QAP)^{-} Q, \qquad (5.5.11)$$

其中 $(QAP)^{-}$ 为任一个 $(QAP)^{(1)}$，则

$$X \in A\{2\}，且\ \mathscr{M}(X) = \mathscr{M}(P),\mathscr{N}(X) = \mathscr{N}(Q)$$

$$\Longleftrightarrow r(QAP) = r(P) = r(Q).$$

注 推论5.5.1表明，当 $r(QAP)=r(P)=r(Q)$ 时，(5.5.11)就是一个具有给定列空间 $\mathscr{M}(P)$ 和零空间 $\mathscr{N}(Q)$ 的{2}-逆.

下面将证明这样的{2}-逆是唯一的，我们的结论将更一般些.

定理5.5.4 对给定的秩为 r 的矩阵 $A\in C^{m\times n}$ 和两个子空间

$$T \subset C^{n},S \subset C^{m},\dim T = t \leqslant r,\dim S = m - t ,$$

则存在满足条件

$$\mathscr{M}(X) = T,\mathscr{N}(X) = S \qquad (5.5.12)$$

的{2}-逆 X 的充要条件为

$$AT \oplus S = C^{m}, \qquad (5.5.13)$$

此时，这样的 X 也是唯一的，这里 AT 定义为

$$AT = \{At,t \in T\}.$$

证明 $\Leftarrow$ 设 $P\in C^{n\times t}$ 秩为 t，其列构成 T 的一组基，于是 $\mathscr{M}(P)=T$. 设 $Q^{*}\in C^{m\times t}$ 秩为 t，其列构成 $S^{\perp}$ 的一组基，于是 $\mathscr{M}(Q^{*})=S^{\perp}$. 所以，$AP$ 的列张成子空间 AT，由(5.5.13)得

$$r(AP) = t. \qquad (5.5.14)$$

另一方面，从(5.5.13)知，

$$AT \cap S = \{0\}, \qquad (5.5.15)$$

根据(5.5.14)和(5.5.15),我们可证明:QAP 可逆.事实上,若 $QAPy=0$,则 $APy\perp S^{\perp}$,因而,$APy\in S$,但 $APy\in AT$,于是从(5.5.15)知 $APy=0$. 结合(5.5.14),便有 $y=0$. 这就证明了 QAP 是可逆阵.

因为 $r(QAP)=r(P)=r(Q)=t$,依推论5.5.1,

$$X = P(QAP)^{-1}Q \in A\{2\}$$

且

$$\mathscr{M}(X) = \mathscr{M}(P) = T,$$

$$\mathscr{N}(X) = \mathscr{N}(Q) = S.$$

充分性得证.

$\Rightarrow$ 设 $X\in A\{2\}$,且满足(5.5.12),则 $A\in X\{1\}$且 AX 为幂等阵.因而,

$$\mathscr{M}(AX) \oplus \mathscr{N}(AX) = C^m, \tag{5.5.16}$$

由 $\mathscr{M}(X)=T$,可推得

$$\mathscr{M}(AX) = A\mathscr{M}(X) = AT,$$

$$\mathscr{N}(AX) = \mathscr{N}(X) = S.$$

从(5.5.16)便得到(5.5.13).必要性得证.

最后证明唯一性.设 X_1和 X_2都是 A 的具有列空间 T 和零空间 S 的{2}-逆,则 $A\in X_1\{1\}$,$A\in X_2\{1\}$,利用(3.6.1)、(3.6.2)及假设条件可推得

$$X_1A = P_{\mathscr{M}(X_1A),\mathscr{N}(X_1A)} = P_{T,\mathscr{N}(X_1A)},$$

$$AX_2 = P_{\mathscr{M}(AX_2),\mathscr{N}(AX_2)} = P_{\mathscr{M}(AX_2),S}.$$

注意到,$\mathscr{M}(X_2)=T$,$\mathscr{N}(X_1)=S$,利用引理3.6.1,

$$X_2 = P_{T,\mathscr{N}(X_1A)}X_2 = (X_1A)X_2 = X_1(AX_2)$$

$$= X_1P_{\mathscr{M}(AX_2),S} = X_1.$$

定理证毕.

注 根据定理5.5.4,满足(5.5.12)的 $A^{(2)}$是唯一的.我们把它记为$A_{T,S}^{(2)}$.

下面的推论是推论5.5.1的另一种表述.

推论5.5.2 使用推论5.5.1的记号,则

$$X = A_{\mathcal{R}(P),\mathcal{N}(Q)}^{(2)} \Longleftrightarrow r(QAP) = r(P) = r(Q).$$

第六章　分块矩阵的广义逆

本章讨论几种常见的分块矩阵的重要广义逆公式，其中包括形如$(A' \vdots B')'$的行分块阵、形如$(A \vdots B)$的列分块阵、四块矩阵以及它的特殊形式镶边矩阵的广义逆. 这些结果在广义逆的理论和应用上都占有重要的地位.

§6.1　行分块矩阵

设$\boldsymbol{A}$和$\boldsymbol{B}$为具有列数为n的矩阵. 考虑两个相容线性方程组

$$\boldsymbol{Ax} = \boldsymbol{a}, \qquad (6.1.1)$$

$$\boldsymbol{Bx} = \boldsymbol{b}. \qquad (6.1.2)$$

根据定理 4.8.1，它们的通解分别为

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{a} + (\boldsymbol{I}_n - \boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{t}, \quad \boldsymbol{t} \in \boldsymbol{C}^n$$

和

$$\boldsymbol{x} = \boldsymbol{B}^{+}\boldsymbol{b} + (\boldsymbol{I}_n - \boldsymbol{B}^{+}\boldsymbol{B})\boldsymbol{t}, \quad \boldsymbol{t} \in \boldsymbol{C}^n.$$

写成流形的形式，分别为

$$\boldsymbol{A}^{+}\boldsymbol{a} + \mathcal{N}(\boldsymbol{A}), \qquad (6.1.3)$$

$$\boldsymbol{B}^{+}\boldsymbol{a} + \mathcal{N}(\boldsymbol{B}) \qquad (6.1.4)$$

（注：关于流形的定义，参见§4.5）.

引理 6.1.1　(6.1.3)和(6.1.4)都是正交表示.

证明　由定理 4.2.5 知

$$\mathcal{M}(\boldsymbol{A}^{+}) = \mathcal{M}(\boldsymbol{A}^{+}\boldsymbol{A}) = (\mathcal{N}(\boldsymbol{A}^{+}\boldsymbol{A}))^{\perp} = \mathcal{N}(\boldsymbol{A})^{\perp}$$

$$= \mathcal{M}(\boldsymbol{P}_{\mathcal{N}(\boldsymbol{A})^{\perp}}),$$

因此

$$\boldsymbol{x}_0 = \boldsymbol{A}^+ \boldsymbol{a} \in \mathscr{N}(\boldsymbol{A})^{\perp},$$

即 $\boldsymbol{x}_0 \perp \mathscr{N}(\boldsymbol{A})$. 这就证明了(6.1.3)是正交表示. 同法可证，(6.1.4)是正交表示.

易见，如果(6.1.1)和(6.1.2)有公共解，则它们的公共解为

$$\{\boldsymbol{A}^+ \boldsymbol{a} + \mathscr{N}(\boldsymbol{A})\} \cap \{\boldsymbol{B}^+ \boldsymbol{b} + \mathscr{N}(\boldsymbol{B})\}, \tag{6.1.5}$$

它也是分块线性方程组

$$\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \tag{6.1.6}$$

的解集. 为了研究这个解集，就必须考虑行分块阵 $\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix}$ 的各种广义逆.

引理 6.1.2 假设方程组(6.1.1)和(6.1.2)相容，则它们有公共解的充要条件为

$$\boldsymbol{A}^+ \boldsymbol{a} - \boldsymbol{B}^+ \boldsymbol{b} \in \mathscr{N}(\boldsymbol{A}) + \mathscr{N}(\boldsymbol{B}). \tag{6.1.7}$$

若(6.1.7)成立，记公共解集为 S，则

(a) $S = \boldsymbol{A}^+ \boldsymbol{a} + \boldsymbol{P}_{\mathscr{N}(\boldsymbol{A})}(\boldsymbol{P}_{\mathscr{N}(\boldsymbol{A})} + \boldsymbol{P}_{\mathscr{N}(\boldsymbol{B})})^+ (\boldsymbol{B}^- \boldsymbol{b} - \boldsymbol{A}^+ \boldsymbol{a})$
$\quad + \mathscr{N}(\boldsymbol{A}) \cap \mathscr{N}(\boldsymbol{B})$;

(b) $S = \boldsymbol{B}^+ \boldsymbol{b} - \boldsymbol{P}_{\mathscr{N}(\boldsymbol{B})}(\boldsymbol{P}_{\mathscr{N}(\boldsymbol{A})} + \boldsymbol{P}_{\mathscr{N}(\boldsymbol{B})})^+ (\boldsymbol{B}^+ \boldsymbol{b} - \boldsymbol{A}^+ \boldsymbol{a})$
$\quad + \mathscr{N}(\boldsymbol{A}) \cap \mathscr{N}(\boldsymbol{B})$;

(c) $S = (\boldsymbol{A}^+ \boldsymbol{A} + \boldsymbol{B}^+ \boldsymbol{B})^- (\boldsymbol{A}^+ \boldsymbol{a} + \boldsymbol{B}^+ \boldsymbol{b})$
$\quad + \mathscr{N}(\boldsymbol{A}) \cap \mathscr{N}(\boldsymbol{B})$,

并且它们都是正交表示.

证明 应用定理 4.5.3，方程组(6.1.1)和(6.1.2)的公共解集(6.1.5)非空当且仅当(6.1.7)成立. 而(a)和(b)分别是定理 4.5.3(a)和推论 4.5.1(a).

下证结论(c). 从(4.2.5)和定理 4.2.5 知

$$A^{+}A = P_{\mathscr{M}(A^{+}A)},$$

$$\mathscr{N}(A) = \mathscr{N}(A^{+}A) = \mathscr{M}(A^{+}A)^{\perp},$$

所以

$$A^{+}A = P_{\mathscr{N}(A)^{\perp}}.$$

同理,

$$B^{+}B = P_{\mathscr{N}(B)^{\perp}}.$$

再应用定理 4.5.3(b),得

$$\begin{aligned}
&\{A^{+}a + \mathscr{N}(A)\} \cap \{B^{+}b + \mathscr{N}(B)\} \\
&= A^{+}a + (A^{+}A + B^{+}B)^{+}B^{+}B(B^{+}b - A^{+}a) \\
&\quad + \mathscr{N}(A) \cap \mathscr{N}(B) \\
&= (A^{+} - (A^{+}A + B^{+}B)^{+}B^{+}BA^{+})a \\
&\quad + (A^{+}A + B^{+}B)^{+}B^{+}b \\
&\quad + \mathscr{N}(A) \cap \mathscr{N}(B).
\end{aligned} \tag{6.1.8}$$

因为

$$\mathscr{M}(A^{+}) = \mathscr{M}(A^{*}) \subset \mathscr{M}(A^{*}) + \mathscr{M}(B^{*}),$$

故有

$$(A^{+}A + B^{+}B)^{+}(A^{+}A + B^{+}B)A^{+} = A^{+}.$$

最后得到

$$\begin{aligned}
&(A^{+} - (A^{+}A + B^{+}B)^{+}B^{+}BA^{+})a \\
&= (A^{+}A + B^{+}B)^{+}[(A^{+}A + B^{+}B)A^{+} - B^{+}BA^{+}]a \\
&= (A^{+}A + B^{+}B)^{+}A^{+}a.
\end{aligned}$$

代入(6.1.8)即得(c). 至于这些表示的正交性质是引理 6.1.1 和推论 4.5.2 的直接结果. 引理证毕.

定理 6.1.1 设矩阵 A 和 B 有相同列数. 记

$$X_1 = (A^{+} \;\vdots\; 0) + P_{\mathscr{N}(A)}(P_{\mathscr{N}(A)} + P_{\mathscr{N}(B)})^{+}(-A^{+} \;\vdots\; B^{+}), \tag{6.1.9}$$

$$X_2 = (0 \vdots B^+) - P_{\mathscr{N}(B)}(P_{\mathscr{N}(A)} + P_{\mathscr{N}(B)})^+ (-A^+ \vdots B^+), \tag{6.1.10}$$

$$X_3 = (A^+ A + B^+ B)^+ (A^+ \vdots B^+), \tag{6.1.11}$$

则

$$X_i \in \begin{pmatrix} A \\ B \end{pmatrix} \{1,2,4\}, \quad i = 1,2,3. \tag{6.1.12}$$

进一步，若假设

$$\mathscr{M}(A^*) \cap \mathscr{M}(B^*) = \{0\}, \tag{6.1.13}$$

则

$$X_1 = X_2 = X_3 = \begin{pmatrix} A \\ B \end{pmatrix}^+.$$

证明 定理的证明分如下三部分：

(a) 先证明(6.1.12). 若方程组(6.1.6)相容，依引理6.1.2知，

$$X_i \begin{pmatrix} a \\ b \end{pmatrix}, \quad i = 1,2,3$$

都是它的解. 因为引理 6.1.2 给出的都是正交表示，于是

$$X_i \begin{pmatrix} a \\ b \end{pmatrix} \perp (\mathscr{N}(A) \cap \mathscr{N}(B)), \quad i = 1,2,3.$$

注意到

$$\mathscr{N}(A) \cap \mathscr{N}(B) = \mathscr{N}\begin{pmatrix} A \\ B \end{pmatrix},$$

因而

$$X_i \begin{pmatrix} a \\ b \end{pmatrix} \perp \mathscr{N}\begin{pmatrix} A \\ B \end{pmatrix} \quad i = 1,2,3.$$

由此可知，

$$X_i\begin{pmatrix} a \\ b \end{pmatrix}, \qquad i = 1,2,3$$

都是(6.1.6)的解集 S 中最小范数解. 再应用定理 5.3.3 可推出

$$X_i \in \begin{pmatrix} A \\ B \end{pmatrix}\{1,4\}.$$

下面再证明

$$X_i \in \begin{pmatrix} A \\ B \end{pmatrix}\{2\}.$$

因为

$$\begin{aligned} -A^+A + B^+B &= -P_{\mathcal{N}(A)^\perp} + P_{\mathcal{N}(B)^\perp} \\ &= -(I - P_{\mathcal{N}(A)}) + I - P_{\mathcal{N}(B)} \\ &= P_{\mathcal{N}(A)} - P_{\mathcal{N}(B)} \\ &= (P_{\mathcal{N}(A)} + P_{\mathcal{N}(B)}) - 2P_{\mathcal{N}(B)}, \end{aligned}$$

于是

$$\begin{aligned} X_1\begin{pmatrix} A \\ B \end{pmatrix} &= A^+A + P_{\mathcal{N}(A)} \\ &\quad \times (P_{\mathcal{N}(A)} + P_{\mathcal{N}(B)})^+(-A^+A + B^+B) \\ &= A^+A + P_{\mathcal{N}(A)}(P_{\mathcal{N}(A)} + P_{\mathcal{N}(B)})^+ \\ &\quad \times [(P_{\mathcal{N}(A)} + P_{\mathcal{N}(B)}) - 2P_{\mathcal{N}(B)}]. \end{aligned}$$

再应用定理 4.5.1 和定理 4.5.2,

$$\begin{aligned} X_1\begin{pmatrix} A \\ B \end{pmatrix} &= A^+A + P_{\mathcal{N}(A)}[P_{\mathcal{N}(A)+\mathcal{N}(B)} - P_{\mathcal{N}(A)\cap\mathcal{N}(B)}] \\ &= A^+A + P_{\mathcal{N}(A)} - P_{\mathcal{N}(A)\cap\mathcal{N}(B)} \\ &= I - P_{\mathcal{N}(A)\cap\mathcal{N}(B)}, \end{aligned} \tag{6.1.14}$$

这里利用了 $P_{\mathcal{N}(A)}=I-A^+A$.

因为

$$P_{\mathcal{N}(A)\cap\mathcal{N}(B)}A^+=0, \tag{6.1.15}$$

$$P_{\mathcal{N}(A)\cap\mathcal{N}(B)}B^+=0, \tag{6.1.16}$$

于是由(6.1.14)得

$$\begin{aligned} X_1\begin{pmatrix}A\\B\end{pmatrix}X_1 &= X_1 - P_{\mathcal{N}(A)\cap\mathcal{N}(B)}P_{\mathcal{N}(A)} \\ &\quad \times (P_{\mathcal{N}(A)}+P_{\mathcal{N}(B)})^+(-A^+ \vdots B^+) \\ &\stackrel{\mathrm{d}}{=} X_1-\Delta_1. \end{aligned}$$

易见,为了证明

$$X_1\in\begin{pmatrix}A\\B\end{pmatrix}\{2\},$$

只需证明 $\Delta_1=0$.

因为

$$P_{\mathcal{N}(A)\cap\mathcal{N}(B)}=P_{\mathcal{N}(A)\cap\mathcal{N}(B)}P_{\mathcal{N}(A)}=P_{\mathcal{N}(A)\cap\mathcal{N}(B)}P_{\mathcal{N}(B)},$$

故

$$\begin{aligned} \Delta_1 &= \frac{1}{2}P_{\mathcal{N}(A)\cap\mathcal{N}(B)}(P_{\mathcal{N}(A)}+P_{\mathcal{N}(B)}) \\ &\quad \times (P_{\mathcal{N}(A)}+P_{\mathcal{N}(B)})^+(-A^+ \vdots B^+). \end{aligned}$$

利用定理 4.5.1,得

$$\begin{aligned} \Delta_1 &= \frac{1}{2}P_{\mathcal{N}(A)\cap\mathcal{N}(B)}P_{\mathcal{N}(A)+\mathcal{N}(B)}(-A^+ \vdots B^+) \\ &= \frac{1}{2}P_{\mathcal{N}(A)\cap\mathcal{N}(B)}(-A^+ \vdots B^+)=0, \end{aligned}$$

这里利用了 $\mathcal{N}(A)\cap\mathcal{N}(B)\subset\mathcal{N}(A)+\mathcal{N}(B)$以及(6.1.15)和(6.1.16). 至此我们证明了

$$X_1\in\begin{pmatrix}A\\B\end{pmatrix}\{1,2,4\}.$$

类似地，可以证明

$$X_2 \in \begin{pmatrix} A \\ B \end{pmatrix}\{2\},$$

从而证明了

$$X_2 \in \begin{pmatrix} A \\ B \end{pmatrix}\{1,2,4\}.$$

容易证明

$$X_3 \in \begin{pmatrix} A \\ B \end{pmatrix}\{2\}.$$

事实上

$$\begin{aligned} X_3\begin{pmatrix} A \\ B \end{pmatrix}X_3 &= (A^+A + B^+B)^+(A^+A + B^+B) \\ &\quad \times (A^+A + B^+B)^+(A^+ \;\vdots\; B^+) \\ &= (A^+A + B^+B)^+(A^+ \;\vdots\; B^+) = X_3, \end{aligned}$$

这就完成了

$$X_3 \in \begin{pmatrix} A \\ B \end{pmatrix}\{1,2,4\}$$

的证明.

(b) 证明在条件(6.1.13)下

$$X_i = \begin{pmatrix} A \\ B \end{pmatrix}^+, \quad i = 1,2.$$

由前面已证部分，我们只需证

$$X_i \in \begin{pmatrix} A \\ B \end{pmatrix}\{3\}, \quad i = 1,2.$$

首先

$$\begin{aligned} BX_1 &= (BA^+ \;\vdots\; 0) + BP_{\mathcal{N}(A)}(P_{\mathcal{N}(A)} + P_{\mathcal{N}(B)})^+ \\ &\quad \times (-A^+ \;\vdots\; B^+) \\ &\stackrel{d}{=} (BA^+ \;\vdots\; 0) + \Delta_2. \end{aligned} \tag{6.1.17}$$

因为

$$\mathscr{N}(A)=\mathscr{M}(A^*)^{\perp}, \mathscr{N}(B)=\mathscr{M}(B^*)^{\perp}$$

以及对任意两个子空间 L 和 M 有关系：

$$(L^{\perp}\cap M^{\perp})^{\perp}=L+M,$$

于是，在条件(6.1.13)下，我们有

$$\mathscr{N}(A)+\mathscr{N}(B)=(\mathscr{M}(A^*)\cap\mathscr{M}(B^*))^{\perp}=C^n.$$

再由定理 4.5.1

$$(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})^{+}=P_{\mathscr{N}(A)+\mathscr{N}(B)}=I_n, \tag{6.1.18}$$

因而

$$\begin{aligned}\Delta_2&=B(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)}-P_{\mathscr{N}(B)})\\&\quad\times(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})^{+}(-A^{+}\ \vdots\ B^{+})\\&=B(I-P_{\mathscr{N}(B)}(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})^{+})(-A^{+}\ \vdots\ B^{+})\\&=B(-A^{+}\ \vdots\ B^{+})=(-BA^{+}\ \vdots\ BB^{+}),\end{aligned}$$

代入(6.1.17)，得

$$BX_1=(0,BB^{+}).$$

但是，显然有 $AX_1=(AA^{+}\ \vdots\ 0)$，于是我们证明了

$$\begin{pmatrix}A\\B\end{pmatrix}X_1=\begin{pmatrix}AA^{+}&0\\0&BB^{+}\end{pmatrix},$$

即

$$X_1\in\begin{pmatrix}A\\B\end{pmatrix}\{3\}.$$

利用(6.1.18)，得

$$\begin{aligned}X_2-X_1&=(-A^{+}\ \vdots\ B^{+})-(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})\\&\qquad\times(P_{\mathscr{N}(A)}+P_{\mathscr{N}(B)})^{+}(-A^{+}\ \vdots\ B^{+})\\&=0,\end{aligned}$$

这就证明了

$$X_1 = X_2 = \begin{pmatrix} A \\ B \end{pmatrix}^+.$$

(c) 最后证明

$$X_3 = \begin{pmatrix} A \\ B \end{pmatrix}^+.$$

依定理 5.1.7，对任意矩阵 D，在广义逆集 $D\{1,2\}$ 中，D^+ 是唯一的具有列空间 $\mathscr{M}(D^*)$ 和零空间 $\mathscr{N}(D^*)$ 的矩阵，因此，D^+ 也是广义逆集合 $D\{1,2,4\}$ 中唯一具有列空间 $\mathscr{N}(D^*)$ 的矩阵. 前面已经证明了

$$X_3 \in \begin{pmatrix} A \\ B \end{pmatrix}\{1,2,4\},$$

因此，下面只要证明 $\mathscr{N}(X_3) = \mathscr{N}(A^* \vdots B^*)$ 即可. 从已证事实可知

$$r(X_3) = r\begin{pmatrix} A \\ B \end{pmatrix} = r(A^* \vdots B^*),$$

于是，我们也只要证明

$$\mathscr{N}(A^* \vdots B^*) \subset \mathscr{N}(X_3). \qquad (6.1.19)$$

设

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \mathscr{N}(A^* \vdots B^*),$$

于是

$$A^* x + B^* y = 0,$$

结合假设(6.1.13)可推出

$$A^* x = -B^* y = 0. \qquad (6.1.20)$$

因为对任意矩阵 D，$\mathscr{N}(D^+) = \mathscr{N}(D^*)$，因而上式等价于 $A^+ x = B^+ y = 0$，于是

$$X_3 \begin{pmatrix} x \\ y \end{pmatrix} = 0,$$

此即

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{N}(X_3),$$

于是(6.1.19)得证. 定理证毕.

应用定理 6.1.1,我们还可获得按列分块的矩阵$(A \vdots B)$的$\{1,2,3\}$-逆. 事实上,因为

$$(A \vdots B) = \begin{pmatrix} A^* \\ B^* \end{pmatrix}^*,$$

于是将定理 6.1.1 应用于$\begin{pmatrix} A^* \\ B^* \end{pmatrix}$,并注意到

$$X \in A\{i\} \Leftrightarrow X^* \in A^*\{i\}, i = 1,2,$$

$$X \in A\{3\} \Leftrightarrow X^* \in A^*\{4\},$$

$$X \in A\{4\} \Leftrightarrow X^* \in A^*\{3\},$$

便可证明如下推论:

推论 6.1.1 设 A 与 B 有相同的行数,记

$$X_1 = \begin{pmatrix} A^+ \\ 0 \end{pmatrix} - \begin{pmatrix} -A^+ \\ B^+ \end{pmatrix} (P_{\mathcal{N}(A^*)} + P_{\mathcal{N}(B^*)})^+ P_{\mathcal{N}(A^*)},$$

$$X_2 = \begin{pmatrix} 0 \\ B^+ \end{pmatrix} - \begin{pmatrix} -A^+ \\ B^+ \end{pmatrix} (P_{\mathcal{N}(A^*)} + P_{\mathcal{N}(B^*)})^+ P_{\mathcal{N}(B^*)},$$

$$X_3 = \begin{pmatrix} A^+ \\ B^+ \end{pmatrix} (AA^+ + BB^+)^+,$$

则 $X_i \in (A \vdots B)\{1,2,3\}, i=1,2,3$. 进一步若假设

$$\mathcal{M}(A) \cap \mathcal{M}(B) = \{0\},$$

则 $X_i = (A \vdots B)^+, i=1,2,3$.

§6.2 列分块矩阵

设 $A \in C^{m \times p}, B \in C^{m \times q}$. 所谓列分块矩阵是指形如$(A \vdots B)$的

矩阵. 本节讨论它的各种广义逆.

定理 6.2.1 记 $C=(I_n-AA^-)B$，则

$$(A \vdots B)^- = \begin{pmatrix} A^- - A^- BC^- (I_n - AA^-) \\ C^- (I_n - AA^-) \end{pmatrix}. \quad (6.2.1)$$

证明 本定理有两种证法. 其一是直接验证(6.2.1)的右端满足 Penrose 第一方程；其二是下面的构造性证法.

记 $D=P_1(A \vdots B)P_2$，其中

$$P_1 = \begin{pmatrix} A^- \\ I_n \end{pmatrix}, P_2 = \begin{pmatrix} I_p & -A^- B \\ 0 & I_q \end{pmatrix}. \quad (6.2.2)$$

D 可变形为

$$D = \begin{pmatrix} A^- A & A^- C \\ A & C \end{pmatrix}. \quad (6.2.3)$$

注意到 $AA^-C=0$，不难验证

$$\begin{pmatrix} I & 0 \\ -C^- A & C^- \end{pmatrix} \in D\{1\}. \quad (6.2.4)$$

因为 P_1 是列满秩阵，P_2 是可逆阵，利用定理 5.4.2(b)，

$$(A \vdots B)^- = P_2 D^- P_1.$$

利用(6.2.4)直接计算可证明上式右端即为(6.2.1)右端. 定理证毕.

注 从上面的证明过程可以看出，对任意的 A^- 和 C^- (6.2.1)的右端是 $(A \vdots B)^-$，但它们未必能给出 $(A \vdots B)\{1\}$. 因此，更确切地，应表示为

$$\begin{pmatrix} A^- - A^- BC^- (I_n - AA^-) \\ C^- (I_n - AA^-) \end{pmatrix} \in (A \vdots B)\{1\}.$$

但习惯上，很多文献都采用(6.2.1)的记法.

若 $\mathscr{M}(B)\subset\mathscr{M}(A)$，则 $C=0$. 此时(6.2.1)有更简单的形

式.

推论 6.2.1 若 $\mathscr{M}(B)\subset\mathscr{M}(A)$，则

$$(A \vdots B)^- = \begin{pmatrix} A^- \\ 0 \end{pmatrix}.$$

定理 6.2.2

$$(A \vdots B)^{(1,2)} = \begin{pmatrix} A^- AA^- - A^- BC^{(1,2)}(I_n - AA^-) \\ C^{(1,2)}(I_n - AA^-) \end{pmatrix}, \tag{6.2.5}$$

其中 C 同定理 6.2.1.

证明 设 P_1, P_2 和 D 同上一定理，可以验证

$$D^{(1,2)} = \begin{pmatrix} A^- A & 0 \\ -C^{(1,2)}A & C^{(1,2)} \end{pmatrix}. \tag{6.2.6}$$

应用推论 5.4.3，得

$$(A \vdots B)^{(1,2)} = P_2 D^{(1,2)} P_1,$$

将(6.2.6)代入上式便得到(6.2.5). 定理证毕.

类似于推论 6.2.1，我们有如下事实.

推论 6.2.2 若 $\mathscr{M}(B)\subset\mathscr{M}(A)$，则

$$(A \vdots B)^{(1,2)} = \begin{pmatrix} A^- AA^- \\ 0 \end{pmatrix}.$$

定理 6.2.3

$$(A \vdots B)^{(1,2,3)} = \begin{pmatrix} A^{(1,2,3)} - A^{(1,2,3)}BC^{(1,2,3)} \\ C^{(1,2,3)} \end{pmatrix},$$

其中

$$C = (I - AA^{(1,2,3)})B.$$

证明 记

$$P = \begin{pmatrix} I & -A^{(1,2,3)}B \\ 0 & I \end{pmatrix}.$$

定义 $D=(A \vdots B)P$,则 $D=(A \vdots C)$.因为

$$A^*C = A^*B - A^*AA^{(1,2,3)}B$$

$$= A^*B - A^*A^{(1,2,3)*}A^*B = 0,$$

我们有

$$D^*D = \begin{pmatrix} A^*A & 0 \\ 0 & C^*C \end{pmatrix}.$$

取

$$(D^*D)^- = \begin{pmatrix} (A^*A)^- & 0 \\ 0 & (C^*C)^- \end{pmatrix},$$

应用定理 5.4.1,得

$$D^{(1,2,3)} = (D^*D)^- D^* = \begin{pmatrix} A^{(1,2,3)} \\ C^{(1,2,3)} \end{pmatrix}.$$

最后由定理 5.4.3 知

$$(A \vdots B)^{(1,2,3)} = PD^{(1,2,3)} = \begin{pmatrix} A^{(1,2,3)} - A^{(1,2,3)}BC^{(1,2,3)} \\ C^{(1,2,3)} \end{pmatrix}.$$

定理证毕.

和前面两个推论一样,当 $\mathscr{M}(B)\subset\mathscr{M}(A)$时,$C=0$. 故有如下推论.

推论 6.2.3 若 $\mathscr{M}(B)\subset\mathscr{M}(A)$.则

$$(A \vdots B)^{(1,2,3)} = \begin{pmatrix} A^{(1,2,3)} \\ 0 \end{pmatrix}.$$

现在我们讨论$(A \vdots B)$的 Moore-Penrose 广义逆.

定理 6.2.4

$$(A \vdots B)^+ = \begin{pmatrix} A^+ - A^+B(C^+ + K) \\ C^+ + K \end{pmatrix}, \qquad (6.2.7)$$

其中

$$C=(I-AA^+)B, \tag{6.2.8}$$
$$K=(I-C^+C)[I+(I-C^+C)$$
$$\times B^*A^{+*}A^+B(I-C^+C)]^{-1}B^*A^{+*}A^+(I-BC^+).$$

证明 记 $D=(A\vdots B)$. 将 D^+ 分块为

$$D^+=\begin{pmatrix} X \\ Y \end{pmatrix},$$

其中 $X\in C^{p\times n}$, $Y\in C^{(q-p)\times n}$, 于是

$$DD^+=AX+BY, \tag{6.2.9}$$

$$D^+D=\begin{pmatrix} XA & XB \\ YA & YB \end{pmatrix}. \tag{6.2.10}$$

由 $DD^+D=D$, 我们得到

$$DD^+A=A, \tag{6.2.11}$$

$$DD^+B=B. \tag{6.2.12}$$

对(6.2.11)两边求转置和共轭, 然后左乘 $(A^*A)^+$, 利用定理 4.2.2(d), 便有

$$A^+DD^+=A^+. \tag{6.2.13}$$

再从关系 $D^+=D^*(D^+)^*D^+$ (见定理 4.2.2(b))得

$$X=A^*X^*X+A^*Y^*Y.$$

因为 A^+A 为向 $\mathscr{M}(A^*)$ 的正交投影阵, 故 $A^+AA^*=A^*$, 于是

$$A^+AX=A^*X^*X+A^*Y^*Y=X. \tag{6.2.14}$$

将(6.2.9)左乘 A^+, 并利用(6.2.13)和(6.2.14), 得

$$A^+=X+A^+BY,$$

这就证明了 $D^+=(A\vdots B)^+$ 具有如下形式:

$$(A\vdots B)^+=\begin{pmatrix} A^+-A^+BY \\ Y \end{pmatrix}. \tag{6.2.15}$$

剩下的问题是证明 $Y=C^{+}+K$.

由(6.2.15)和(6.2.8)有

$$DD^{+}=AA^{+}+(I-AA^{+})BY=AA^{+}+CY. \tag{6.2.16}$$

因为 $A^{+}C=0$,应用定理 4.2.8 可推出

$$C^{+}A=0. \tag{6.2.17}$$

另外利用(6.2.11)和(6.2.12),我们有

$$DD^{+}C=DD^{+}B-DD^{+}AA^{+}B=B-AA^{+}B=C,$$

等价地, $C^{*}=C^{*}DD^{+}$, 因而 $C^{+}=C^{+}DD^{+}$. 用 C^{+} 左乘(6.2.16)并利用(6.2.17),得到

$$C^{+}CY=C^{+},$$

这表明 Y 满足方程

$$CY=CC^{+}. \tag{6.2.18}$$

另外

$$CC^{+}C=CC^{+}(B-AA^{+}B)=CC^{+}B=C,$$

将最后一个等号两边左乘 C^{+} 得

$$C^{+}B=C^{+}C \tag{6.2.19}$$

再结合(6.2.18)得 $CYB=CC^{+}B=C$. 从(6.2.10)可以看出 YB 是一个 Hermite 阵,于是

$$C^{*}=(YB)^{*}C^{*}=YBC^{*}.$$

所以

$$YBC^{+}=C^{+}. \tag{6.2.20}$$

从 A^{+}的定义和(6.2.15)、(6.2.16)可以得到

$$YAA^{+}+YCC^{+}=Y, \tag{6.2.21}$$

$$A^{+}BYA=(A^{+}BYA)^{*}, \tag{6.2.22}$$

$$(YA)^{*}=A^{+}B(I-YB). \tag{6.2.23}$$

下面我们用这三个方程来证明 $Y=C^{+}+K$. 从 C 的定义和

(6.2.13)有

$$(YA)^* = A^+ B[I - Y(AA^+ B + C)]$$
$$= A^+ B - A^+ BYAA^+ B - A^+ BYC.$$

应用(6.2.17)和(6.2.18),得

$$(I - C^+ C)YA = YA - C^+ CC^+ YA$$
$$= YA - C^+ YA$$
$$= Y,$$
$$(YA)^* = (A^+ B - A^+ BYAA^+ B)(I - C^+ C),$$

于是

$$YA = (I - C^+ C)B^* (A^+)^*$$
$$- (I - C^+ C)B^* (A^+)^* A^+ BYA.$$

在上式右边用$(I-C^+C)YA$代替YA,给出

$$[I + (I - C^+ C)B^+ (A^+)^* A^+ B(I - C^+ C)]YA$$
$$= (I - C^+ C)B^* (A^+)^*.$$

再用$A^+(I-BC^+)$右乘上式,并利用(6.2.19)、(6.2.20)和(6.2.21)得

$$N(Y - C^+) = (I - C^+ C)B^* (A^+)^* A^+ (I - BC^+), \tag{6.2.24}$$

其中

$$N = I + (I - C^+ C)B^* (A^+)^* A^+ B(I - C^+ C),$$

这是一个正定阵. 于是从(6.2.24)解得

$$Y = C^+ + N^{-1}(I - C^+ C)B^* (A^+)^* A^+ (I - BC^+).$$

最后再利用$I-C^+C$与N的可交换性就完成了定理的证明.

这个定理给出的$(A \vdots B)^+$表达式较复杂. 下面的几个推论是若干简单特例.

推论 6.2.4 C的定义同定理6.2.4. 若$A^+BC^- = 0$,则

$$(A \vdots B)^+ = \begin{pmatrix} A^+ - A^+ BM^{-1}B^*(A^+)^*A^+ \\ C^+ - M^{-1}B^*(A^+)^*A^+ \end{pmatrix}, \tag{6.2.25}$$

其中 $M = I + B^*(A^+)^*A^+B$.

推论 6.2.5 若 $\mathscr{M}(B) \subset \mathscr{M}(A)$. 则

$$(A \vdots B)^+ = \begin{pmatrix} A^+ - A^+ BM^{-1}B^*(A^+)^*A^+ \\ M^{-1}B^*(A^+)^*A^+ \end{pmatrix}, \tag{6.2.26}$$

其中 M 的定义同推论 6.2.4.

证明 当 $\mathscr{M}(B) \subset \mathscr{M}(A)$时, $AA^+B = B$, 于是 $C = 0$, $C^+ = 0$. 因此推论 6.2.4 的条件成立, 并且(6.2.25)化为(6.2.26). 证毕.

推论 6.2.6 C 的定义同定理 6.2.4. 若 $BC^+B = B$. 则

$$(A \vdots B)^+ = \begin{pmatrix} A^+ - A^+ BC^+ \\ C^+ \end{pmatrix}. \tag{6.2.27}$$

证明 若 $BC^+B = B$, 利用(6.2.19)得 $B = BC^-C$, 此时 $K = 0$, 于是得到(6.2.27).

推论 6.2.7 若 $C = B$ 或等价地 $\mathscr{M}(B) \subset \mathscr{M}(A)^\perp$. 则

$$(A \vdots B)^+ = \begin{pmatrix} A^+ \\ B^+ \end{pmatrix}. \tag{6.2.28}$$

证明 若 $C = B$, 则 $BC^+B = B$. 于是推论 6.2.6 结论成立. 但此时 $A^+BC^+ = A^+CC^+ = 0$, 这里应用了(6.2.17). 这样一来, (6.2.27)化为(6.2.28). 证毕.

注 推论 6.2.7 的逆命题也成立, 即若(6.2.28)成立, 必有 $C = B$. 事实上, 若(6.2.28)成立, 从 $DD^+D = D$ 得到

$$AA^+B + BB^+B = AA^+B + B = B,$$

于是 $AA^+B = 0$, 这就推出 $C = B$.

推论 6.2.8 若 $BC^+B=B$,其中 $C=B(I-A^+A)$,则

$$\begin{pmatrix} A \\ B \end{pmatrix}^+ = (A^+ - C^+ BA^+ \ \vdots\ C^+).$$

证明 对矩阵

$$\begin{pmatrix} A \\ B \end{pmatrix}^* = (A^* \ \vdots\ B^*)$$

应用推论 6.2.6,再将所得等式两边求转置与共轭便得到所要结论.

下面的推论是定理 6.2.4 中 B 为列向量的特殊情形. 记 $A=(a_1,\cdots,a_n)$,A_k 表示 A 的前 k 列组成的子矩阵,于是 $A=(A_{n-1},a_n)$,

$$A_k = (A_{k-1},a_k), \qquad k = 2,\cdots,n,$$

这里约定 $A_n=A$,$A_1=a_1$,

推论 6.2.9

$$A_k^+ = \begin{pmatrix} A_{k-1}^+ - d_k b_k^* \\ b_k^* \end{pmatrix}, \tag{6.2.29}$$

其中

$$d_k = A_{k-1}^+ a_k,$$
$$c_k = a_k - A_{k-1} d_k,$$
$$b_k^* = \begin{cases} c_k^+, & \text{若 } c_k \neq 0, \\ \dfrac{d_k^* A_{k-1}^+}{1 + d_k^* d_k}, & \text{若 } c_k = 0. \end{cases}$$

证明 对现在的情形,定理 6.2.4 中矩阵 C 就是这里的列向量 c_k. 下面分 $c_k\neq 0$ 和 $c_k=0$ 两种情形来证明 .

(a) 若 $c_k\neq 0$. 由推论 4.1.4 知,$c_k^+c_k=1$. 结合(6.2.19)得到 $c_k^+a_k=1$. 因而 $a_kc_k^+a_k=a_k$. 即推论 6.2.6 的假设条件成立,应用该推论,命题得证.

(b) 若 $c_k=0$，则 $c_k\in\mathcal{M}(A_{k-1})$. 由推论 6.2.5，结论得证. 推论证毕.

这个推论给出了计算 A^+ 的一种递推方法，称为 Greville 法，详细讨论见 § 8.5.

记

$$A=\begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix},$$

用 $A_{(k)}$ 表示 A 的前 k 行，则

$$A_{(k)}=\begin{pmatrix} A_{(k-1)} \\ a_k^* \end{pmatrix}.$$

对 $A_{(k)}^*=(A_{(k-1)}^*, a_k)$ 应用推论 6.2.9，求得 $A_{(k)}^{*+}$ 之后，求转置与共轭，便得到如下推论.

推论 6.2.10

$$A_{(k)}^+=(A_{(k-1)}^+-b_k d_k^* \ \vdots\ b_k),$$

其中

$$d_{(k)}^*=a_k^* A_{(k-1)}^+,$$

$$c_k^*=a_k^*-d_k^* A_{(k-1)},$$

$$b_k=\begin{cases} c_k^{*+}, & 若\ c_k^*\neq 0, \\ \dfrac{A_{(k-1)}^+ d_k}{1+d_k^* d_k}, & 若\ c_k^*=0. \end{cases}$$

最后我们给出分块矩阵 $(A \ \vdots\ a)$ 的其他一些广义逆，这些结果可以应用前面的定理来证明或者直接验证.

定理 6.2.5 设 $A\in C^{m\times n}, a\in C^m$. 记

$$X=\begin{pmatrix} G-db^* \\ b^* \end{pmatrix},$$

$$d = Ga.$$

(a) 若 $a \overline{\in} \mathscr{M}(A)$，并且 $b = c/(c^* a)$，其中 $c = (I - AG)^* (I - AG)a$. 则当 $G = A^-$ 时，$X \in (A \vdots a)\{1\}$. 而当 $G = A^{(1,i)}$，$i = 2,3,4$ 时，$X \in (A \vdots a)\{1,i\}$，$i = 2,3,4$.

(b) 若 $a \in \mathscr{M}(A)$，则

(1) 当 $G = A^-$，$b \in C^m$ 时，$X \in (A \vdots a)\{1\}$；

(2) 当 $G = A^{(1,2)}$，$b \in \mathscr{M}(G^*)$ 时，$X \in (A \vdots a)\{1,2\}$；

(3) 当 $G = A^{(1,3)}$，$b \in C^m$ 时，$X \in (A \vdots a)\{1,3\}$；

(4) 当 $G = A^{(1,4)}$，$b = G^* Ga/(1 + a^* G^* Ga)$ 时，$X \in (A \vdots a)\{1,4\}$.

§6.3 四块矩阵

在分块矩阵中，形如

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{6.3.1}$$

的四块矩阵在很多领域都具有重要应用. 本节讨论这种分块矩阵的各种广义逆.

我们先讨论简单而重要的情况. 假定 A 是准对角阵，即 $A_{12} = 0, A_{21} = 0$，也就是

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}. \tag{6.3.2}$$

将 A^- 也做适当分块：

$$A^- = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

从方程 $AA^-A = A$，我们得到

$$A_{11}B_{11}A_{11} = A_{11},$$

$$A_{22}B_{22}A_{22} = A_{22},$$
$$A_{11}B_{12}A_{22} = 0,$$
$$A_{22}B_{21}A_{11} = 0.$$

从前两个方程，立得

$$B_{11} = A_{11}^{-}, \quad B_{22} = A_{22}^{-}.$$

应用定理 3.3.1，后两个方程的解分别为

$$B_{12} = Y - A_{11}^{-}A_{11}YA_{22}A_{22}^{-},$$
$$B_{21} = Z - A_{22}^{-}A_{22}ZA_{11}A_{11}^{-},$$

其中 Y 和 Z 为适当阶数的任意阵. 于是

$$A^{-} = \begin{pmatrix} A_{11}^{-} & Y - A_{11}^{-}A_{11}YA_{22}A_{22}^{-} \\ Z - A_{22}^{-}A_{22}ZA_{11}A_{11}^{-} & A_{22}^{-} \end{pmatrix}, \tag{6.3.3}$$

这里当 A_{11}^{-}，A_{22}^{-}，Y 和 Z 变化时，右端给出了 A^{-} 的全体，即 $A\{1\}$. 特别，当取 $Y=0$，$Z=0$ 时，我们得到

$$A^{-} = \begin{pmatrix} A_{11}^{-} & 0 \\ 0 & A_{22}^{-} \end{pmatrix}, \tag{6.3.4}$$

这时，当 A_{11}^{-}和 A_{22}^{-}变化时，上式右端只给出了 $A\{1\}$的一个子集. 尽管如此，由于(6.3.4)结构简单，加之在许多应用上，我们并不必求出全部的广义逆，因此它仍具有广泛的应用.

现在回到一般情况，即 A_{12}和 A_{21}不必为零阵. 因为研究四块矩阵(6.3.1)的普通逆矩阵的思路和处理技巧可以直接应用到广义逆的情况，所以我们的讨论从 A 可逆的情形开始.

记

$$A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21},$$
$$A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12},$$

它们分别是 A_{22} 和 A_{11} 在 A 中的 Schur 补(见 § 2.5).

定理 6.3.1 假设 A 可逆且分块为(6.3.1).

(a) 若 $\det A_{11}\neq 0$,则 $\det A_{22.1}\neq 0$,且

$$A^{-1}=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix} \tag{6.3.5}$$

(b) 若 $\det A_{22}\neq 0$,则 $\det A_{11.2}\neq 0$,且

$$A^{-1}=\begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}. \tag{6.3.6}$$

证明 因为(6.3.5)和(6.3.6)的证明完全类似,所以我们只证明前者.

记

$$P=\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix},\ Q=\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix},$$

则

$$PAQ=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix} \stackrel{\mathrm{d}}{=} D. \tag{6.3.7}$$

由此可推得 $\det A_{22.1}\neq 0$. 将上式两端求逆矩阵,整理后得

$$\begin{aligned}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} &= QD^{-1}P \\ &= \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^{-1} \end{pmatrix} \\ &\quad\times\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}. \end{aligned} \tag{6.3.8}$$

将右端三个矩阵相乘便得到(6.3.5). 证毕.

当 A 不可逆时,即使 A_{11} 或 A_{22} 可逆,$A_{11.2}$ 或 $A_{22.1}$ 是不可逆

的. 这时在(6.3.5)和(6.3.6)中将 A^{-1}改为 A^{-}, $A_{22\cdot1}^{-1}$和 $A_{11\cdot2}^{-1}$分别改为 $A_{22\cdot1}^{-}$和 $A_{11\cdot2}^{-}$,所得关系式仍成立. 此即

定理 6.3.2 假设 A 有分块(6.3.1).

(a) 若 A_{11}^{-1}存在,则

$$A^{-}=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22\cdot1}^{-}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot1}^{-} \\ -A_{22\cdot1}^{-}A_{21}A_{11}^{-1} & A_{22\cdot1}^{-} \end{pmatrix}. \tag{6.3.9}$$

(b) 若 A_{22}^{-1}存在,则

$$A^{-}=\begin{pmatrix} A_{11\cdot2}^{-} & -A_{11\cdot2}^{-}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot2}^{-} & A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11\cdot2}^{-}A_{12}A_{22}^{-1} \end{pmatrix}. \tag{6.3.10}$$

证明 (a) 首先注意到,当 A_{11}^{-1}存在时,(6.3.7)仍然成立. 于是

$$A^{-}=QD^{-}P, \tag{6.3.11}$$

将

$$D^{-}=\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22\cdot1}^{-} \end{pmatrix}$$

代入(6.3.11)即得(6.3.9).

(b) 证明与(a)相类似,从略. 定理证毕.

在前面的讨论中,不管 A 是否可逆,我们总是假定 A_{11}^{-1}或 A_{22}^{-1}存在. 然而当 $A\geqslant 0$(即 A 为半正定 Hermite 阵)时,这些条件可以去掉. 为了证明相应的结果,我们先证明如下引理.

引理 6.3.1 设 A 分块为(6.3.1)且 $A\geqslant 0$. 则矩阵 $A_{21}A_{11}^{-}A_{12}$和 $A_{12}A_{22}^{-}A_{21}$都与 A_{11}^{-}、A_{22}^{-}选择无关.

证明 因 $A\geqslant 0$,故存在矩阵 $B=(B_1 \vdots B_2)$,使得 $A=B^{*}B$,即

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} B_1^* B_1 & B_1^* B_2 \\ B_2^* B_1 & B_2^* B_2 \end{pmatrix}. \tag{6.3.12}$$

于是

$$A_{21}A_{11}^- A_{12} = B_2^* B_1 (B_1^* B_1)^- B_1^* B_2,$$

$$A_{12}A_{22}^- A_{21} = B_1^* B_2 (B_2^* B_2)^- B_2^* B_1,$$

依据定理 3.2.5(a),上面两式右端都与所含广义逆的选择无关.引理证毕.

定理 6.3.3 设 A 分块为(6.3.1)且 $A \geqslant 0$,则

$$A^- = \begin{pmatrix} A_{11}^- + A_{11}^- A_{12} A_{22.1}^- A_{21} A_{11}^- & -A_{11}^- A_{12} A_{22.1}^- \\ -A_{22.1}^- A_{21} A_{11}^- & A_{22.1}^- \end{pmatrix}, \tag{6.3.13}$$

等价地

$$A^- = \begin{pmatrix} A_{11.2}^- & -A_{11.2}^- A_{12} A_{22}^- \\ -A_{22}^- A_{21} A_{11.2}^- & A_{22}^- + A_{22}^- A_{21} A_{11.2}^- A_{12} A_{22}^- \end{pmatrix}, \tag{6.3.14}$$

其中 $A_{11.2} = A_{11} - A_{12}A_{22}^- A_{21}$, $A_{22.1} = A_{22} - A_{21}A_{11}^- A_{12}$.

证明 根据引理 6.3.1,A 有表达式(6.3.12).应用定理 3.2.5(b),我们得到

$$A_{21}A_{11}^- A_{11} = B_2^* B_1 (B_1^* B_1)^- B_1^* B_1 = B_2^* B_1 = A_{21}, \tag{6.3.15}$$

$$A_{11}A_{11}^- A_{12} = B_1^* B_1 (B_1^* B_1)^- B_1^* B_2 = B_1^* B_2 = A_{12}, \tag{6.3.16}$$

据此,可以验证

$$PAQ = D, \tag{6.3.17}$$

这里

$$P=\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-} & I \end{pmatrix}, Q=\begin{pmatrix} I & -A_{11}^{-}A_{12} \\ 0 & I \end{pmatrix}, \tag{6.3.18}$$

$$D=\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}. \tag{6.3.19}$$

显然,P 和 Q 为可逆阵. 从(6.3.11)得

$$A^{-}=QD^{-}P=Q\begin{pmatrix} A_{11}^{-} & 0 \\ 0 & A_{22.1}^{-} \end{pmatrix}P. \tag{6.3.20}$$

这就证明了(6.3.13). 用类似方法可以证明(6.3.14). 定理证毕.

下面的定理给出了 $A^{(1,2)}$.

定理 6.3.4 设 A 分块为(6.3.1),且 $A\geqslant 0$. 则

$$A^{(1,2)}=\begin{pmatrix} A_{11}^{(1,2)}+A_{11}^{-}A_{12}A_{22.1}^{(1,2)}A_{21}A_{11}^{-} & -A_{11}^{-}A_{12}A_{22.1}^{(1,2)} \\ -A_{22.1}^{(1,2)}A_{21}A_{11}^{-} & A_{22.1}^{(1,2)} \end{pmatrix}, \tag{6.3.21}$$

或

$$A^{(1,2)}=\begin{pmatrix} A_{11.2}^{(1,2)} & -A_{11.2}^{(1,2)}A_{12}A_{22}^{-} \\ -A_{22}^{-}A_{21}A_{11.2}^{(1,2)} & A_{22}^{(1,2)}+A_{22}^{-}A_{21}A_{11.2}^{(1,2)}A_{12}A_{22}^{-} \end{pmatrix}. \tag{6.3.22}$$

证明 采用定理 6.3.3 中的记号,因为

$$D^{(1,2)}=\begin{pmatrix} A_{11}^{(1,2)} & 0 \\ 0 & A_{22.1}^{(1,2)} \end{pmatrix},$$

对(6.3.17)应用定理 5.1.4 得

$$A^{(1,2)}=(P^{-1}DQ^{-1})^{(1,2)}=QD^{(1,2)}P,$$

这就证明了(6.3.21),同理可证(6.3.22). 定理证毕.

定理 6.3.5 设 $A\geqslant 0$ 且分块为(6.3.1),满足条件

$$r(A_{11})+r(A_{22})=r(A), \tag{6.3.23}$$

则

$$A^{(1,2,3)} = \begin{pmatrix} A_{11}^{(1,2,3)} + A_{11}^{-}A_{12}A_{22.1}^{(1,2,3)}A_{21}A_{11}^{-} & -A_{11}^{-}A_{12}A_{22.1}^{(1,2,3)} \\ -A_{22.1}^{(1,2,3)}A_{21}A_{11}^{-} & A_{22.1}^{(1,2,3)} \end{pmatrix}, \tag{6.3.24}$$

或

$$A^{(1,2,3)} = \begin{pmatrix} A_{11.2}^{(1,2,3)} & -A_{11.2}^{(1,2,3)}A_{12}A_{22}^{-} \\ -A_{22}^{-}A_{21}A_{11.2}^{(1,2,3)} & A_{22}^{(1,2,3)} + A_{22}^{-}A_{21}A_{11.2}^{(1,2,3)}A_{12}A_{22}^{-} \end{pmatrix}, \tag{6.3.25}$$

证明 利用定理 6.3.3 及(6.3.15)、(6.3.16)可以证明

$$AA^{-} = \begin{pmatrix} A_{11}A_{11}^{-} & 0 \\ (I - A_{22.1}A_{22.1}^{-})A_{21}A_{11}^{-} & A_{22.1}A_{22.1}^{-} \end{pmatrix}. \tag{6.3.26}$$

为了获得 $A^{(1,2,3)}$，我们要求 A^{-} 满足条件：AA^{-} 为 Hermite 阵. 等价地，

(a) $A_{11}A_{11}^{-}$ 和 $A_{22.1}A_{22.1}^{-}$ 为 Hermite 阵. (6.3.27)

(b) $(I-A_{22.1}A_{22.1}^{-})A_{21}A_{11}^{-}=0$. (6.3.28)

采用(6.3.18)、(6.3.19)的记号，并记

$$D^{(1,2,3)} = \begin{pmatrix} A_{11}^{(1,2,3)} & 0 \\ 0 & A_{22.1}^{(1,2,3)} \end{pmatrix}.$$

因为 $A=P^{-1}DQ^{-1}$，则

$$A^{(1,2,3)} = QD^{(1,2,3)}P,$$

显然(6.3.27)满足，下面我们证明，(6.3.23)等价于(6.3.28).

我们先证：(6.3.28)等价于

$$r(A_{22.1}) = r(A_{22}) \tag{6.3.29}$$

因为 $A\geqslant 0$，故存在矩阵 $B=(B_1 \vdots B_2)$，使得 $A=B'B$，于是(6.3.12)成立，从关系

$$A_{22.1} = A_{22} - A_{21}A_{11}^{-}A_{12} = B_2^{*}B_2 - B_2^{*}B_1(B_1^{*}B_1)^{-}B_1^{*}B_2,$$

可推出

$$\mathscr{M}(A_{22.1}) \subset \mathscr{M}(B_2^{*}) = \mathscr{M}(B_2^{*}B_2) = \mathscr{M}(A_{22}) \tag{6.3.30}$$

若(6.3.29)成立，则 $\mathscr{M}(A_{22.1})=\mathscr{M}(A_{22})$. 于是存在矩阵 E，使得 $B_2^{*}=A_{22.1}E$. 故有

$$A_{22.1}A_{22.1}^{-}A_{21} = A_{22.1}A_{22.1}^{-}B_2^{*}B_1 = A_{22.1}EB_1 = B_2^{*}B_1 = A_{21},$$

这就证明了(6.3.28).

反过来，若(6.3.28)成立，用 A_{12} 右乘之，化简可得到

$$A_{22} - A_{22.1}A_{22.1}^{-}A_{22} = 0.$$

因而

$$\mathscr{M}(A_{22}) \subset \mathscr{M}(A_{22.1}),$$

这表明 $\mathscr{M}(A_{22})=\mathscr{M}(A_{22.1})$. 于是(6.3.29)成立.

最后证明(6.3.23)⇔(6.3.29). 事实上，因为 P 和 Q 皆可逆，故

$$r(A) = r(D) = r(A_{11}) + r(A_{22.1}).$$

因此，结论得证. 定理证毕.

完全类似地，我们有

定理 6.3.6 设 $A\geqslant 0$ 且分块为(6.3.1). 若 $r(A)=r(A_{11})+r(A_{22})$，则

$$A^{+} = \begin{pmatrix} A_{11}^{+} + A_{11}^{+}A_{12}A_{22.1}^{+}A_{21}A_{11}^{+} & -A_{11}^{+}A_{12}A_{22.1}^{+} \\ -A_{22.1}^{+}A_{21}A_{11}^{+} & A_{22.1}^{+} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11.2}^{+} & -A_{11.2}^{+}A_{12}A_{22}^{+} \\ -A_{22}^{+}A_{21}A_{11.2}^{+} & A_{22}^{+} + A_{22}^{+}A_{21}A_{11.2}^{+}A_{12}A_{22}^{+} \end{pmatrix}.$$

§6.4 镶边矩阵

形如

$$M=\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix} \tag{6.4.1}$$

的矩阵称为镶边矩阵((bordered matrix). 显然,这是一种特殊形式的四块矩阵,因此上节的基本方法可以应用到这种情况. 例如,为了求 M^-,我们先求两个可逆阵 P_1 和 P_2,使得

$$P_1MP_2=D=\begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \tag{6.4.2}$$

则

$$M^-=P_2D^-P_1=P_2\begin{pmatrix} D_1^- & 0 \\ 0 & D_2^- \end{pmatrix}P_1.$$

至于其它的广义逆,可以作同样的处理,但需要做一些必要的修正.

定理 6.4.1 设 $S\in C^{n\times n}, S\geqslant 0, L\in C^{m\times n}$. 则

$$\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}^-=\begin{pmatrix} T^- - T^-L^*Q^-LT^- & T^-L^*Q^- \\ Q^-LT^- & Q^-Q-Q^- \end{pmatrix}, \tag{6.4.3}$$

这里 $T=S+L^*L, Q=LT^-L^*$.

证明 令

$$F=\begin{pmatrix} I & L^* \\ 0 & I \end{pmatrix},$$

则

$$FM = \begin{pmatrix} T & L^* \\ L & 0 \end{pmatrix}.$$

因为 $S\geqslant 0$，故存在矩阵 K，使得 $S=K^*K$. 因而

$$T = K^*K + L^*L = (K^* \vdots L^*)\begin{pmatrix} K \\ L \end{pmatrix},$$

这蕴含着

$$\mathscr{M}(L^*) \subset \mathscr{M}(T), \tag{6.4.4}$$

于是可以证明

$$L - LT^- T = 0, \tag{6.4.5}$$

$$L^* - TT^- L^* = 0. \tag{6.4.6}$$

应用这两个关系不难验证

$$\begin{pmatrix} I & 0 \\ -LT^- & I \end{pmatrix}\begin{pmatrix} T & L^* \\ L & 0 \end{pmatrix}\begin{pmatrix} I & -T^- L^* \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} T & 0 \\ 0 & -LT^- L^* \end{pmatrix}. \tag{6.4.7}$$

若记

$$P_1 = \begin{pmatrix} I & 0 \\ -LT^- & I \end{pmatrix} F,$$

$$P_2 = \begin{pmatrix} I & -T^- L^* \\ 0 & I \end{pmatrix},$$

则(6.4.7)变形为

$$P_1\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}P_2 = \begin{pmatrix} T & 0 \\ 0 & -Q \end{pmatrix}.$$

因为 P_1 和 P_2 皆可逆，于是

$$M^- = P_2\begin{pmatrix} T^- & 0 \\ 0 & -Q^- \end{pmatrix}P_1$$

$$= \begin{pmatrix} T^- - T^- L^* Q^- LT^- & T^- L^* Q^- + T^- L^* (I - Q^- Q) \\ Q^- LT^- & Q^- Q - Q^- \end{pmatrix}. \tag{6.4.8}$$

由(6.4.4)知,Q 与所含的广义逆 T^- 的选择无关,若特别取可逆的 T^-,则可推出 $\mathscr{M}(L) \subset \mathscr{M}(Q)$,故有

$$L^* = L^* Q^- Q.$$

代入(6.4.8),便得到 6.4.3).证毕.

分析一下定理的证明过程是有益的.为了找出两个可逆阵 P_1 和 P_2,使得(6.4.2)成立,我们先将 M 左乘了一个 F.F 的作用是将 M 变成

$$\begin{pmatrix} T & L^* \\ L & 0 \end{pmatrix},$$

这个矩阵的特点是,关系式(6.4.4)成立,因而有(6.4.5)和(6.4.6),这两个关系保证了将 M 分别左乘 P_1 和右乘 P_2 后变成对角阵.因此,如果我们假定 $\mathscr{M}(L^*) \subset \mathscr{M}(S)$,则不需要引进矩阵 F.只需令

$$P_1 = \begin{pmatrix} I & 0 \\ -LT^- & I \end{pmatrix},$$

P_2 的定义保持不变.用定理的证明方法可得到如下推论.

推论 6.4.1 设 $S \in C^{n \times n}$,$S \geqslant 0$ 时,$L \in C^{p \times n}$,且 $\mathscr{M}(L^*) \subset \mathscr{M}(S)$,则

$$\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}^- = \begin{pmatrix} S^- - S^- L^* H^- LS^- & S^- L^* H^- \\ H^- LS^- & -H^- \end{pmatrix},$$

这里 $H = LS^- L^*$.

推论 6.4.2 设 $S \in C^{n \times n}$,$S \geqslant 0$,$L \in C^{p \times n}$,$\mathscr{M}(L^*) \cap \mathscr{M}(S) = \{0\}$,$r(L) + r(S) = n$,$r(L) = p$,则

(a) $T=S+L^*L$ 和 $\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}$ 都是可逆阵.

(b) $\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}^{-1}=\begin{pmatrix} T^{-1}ST^{-1} & T^{-}L^* \\ LT^{-1} & 0 \end{pmatrix}$.

证明 (a) 因为 $S\geqslant 0$,故存在 $K\in C^{n\times n}$,使得 $S=K^*K$,故

$$T=(K^*\ \vdots\ L^*)\begin{pmatrix} K \\ L \end{pmatrix}.$$

因而 $\mathscr{M}(T)=\mathscr{M}(K^*\ \vdots\ L^*)$. 依假设 $\mathscr{M}(L^*)\cap\mathscr{M}(S)=\{0\}$,可推出 $\mathscr{M}(L^*)\cap\mathscr{M}(K^*)=\{0\}$,最后得

$$\begin{aligned} r(T)&=r(K^*\ \vdots\ L^*)=r(K^*)+r(L^*) \\ &=r(S)+r(L)=n. \end{aligned}$$

这就证明了 T 可逆. 另一方面

$$\begin{aligned} r\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}&=r\begin{pmatrix} S \\ L \end{pmatrix}+r\begin{pmatrix} L^* \\ 0 \end{pmatrix} \\ &=r(S)+r(L)+r(L^*)=n+p. \end{aligned}$$

这就证明了(a).

(b) 从定理 6.4.1 知,我们只需证明,在现在的假设条件下,$Q=I_p$.

因为 $\mathscr{M}(L^*)\cap\mathscr{M}(S)=\{0\}$,故 S 和 L 的线性无关的行向量构成 C^n 的基. 据此可证明

$$Q=LT^{-}L^*=L(L^*L)^{-}L^*,$$

即 Q 为向 $\mathscr{M}(L)$的正交投影阵. 但 $r(Q)=r(L)=p$,因而 $Q=I_p$. 推论证毕.

现在考虑 $M^{(1,2)}$. 若(6.4.2)成立,从定理 5.1.4 有

$$M^{(1,2)}=P_2\begin{pmatrix} D_1^{(1,2)} & 0 \\ 0 & D_2^{(1,2)} \end{pmatrix}P_1,$$

于是我们证明了如下定理.

定理 6.4.2 在定理 6.4.1 条件下,

$$\begin{pmatrix} S & L^* \\ L & 0 \end{pmatrix}^{(1,2)} = \begin{pmatrix} T^{(1,2)} - T^- L^* Q^{(1,2)} L T^- & T^- L^* Q^{(1,2)} \\ Q^{(1,2)} L T^- & -Q^{(1,2)}(I - Q) \end{pmatrix}, \tag{6.4.9}$$

这里 T 和 Q 的定义同定理 6.4.1.

对于其他的 $\{i, j, \cdots, l\}$-逆,在一般情况下,类似于 (6.4.9) 的结论并不成立.

第七章　广义逆不等式

本章旨在讨论含有广义逆矩阵的一些重要不等式，它们都是矩阵的一种偏序关系.

所谓偏序(partial ordering)，是指在一个集合上定义的一种关系，记为“$>$”，它满足如下三条性质：

(a)　自反性：对一切 $x\in S$，总有 $x>x$ 成立；

(b)　传递性：若 x、y、$z\in S$，且 $x>y$，$y>z$，则必有 $x>z$；

(c)　若 x、$y\in S$，且 $x>y$ 和 $y>x$ 同时成立，则必有 $x=y$.

在定义了偏序的集合中，并不是任何两个元素 x 和 y 之间，总有关系 $x>y$ 或 $y>x$ 成立. 事实上，存在一些元素 x，y，两种关系 $x>y$ 或 $y>x$ 都不成立. 这就是称此关系为 partial ordering 的原因所在.

在矩阵论中有三种偏序，而应用最广泛的是 Löwner 偏序. 本章只讨论含广义逆的 Löwner 偏序，至于一般情况下的 Löwner 偏序以及其它两种偏序读者可参阅王松桂等(1994)及其所引文献.

若 S 为 n 阶半正定 Hermite 阵的全体. 若 $\boldsymbol{A}$、$\boldsymbol{B}\in S$，且满足 $\boldsymbol{A}-\boldsymbol{B}\geqslant 0$，则称在 Löwner 意义下 $\boldsymbol{B}$ 小于 $\boldsymbol{A}$，记为 $\boldsymbol{B}\leqslant\boldsymbol{A}$ 或 $\boldsymbol{A}\geqslant\boldsymbol{B}$. 容易验证，$S$ 上的这种关系是一种偏序，称为 Löwner 偏序. 特别当 $\boldsymbol{A}-\boldsymbol{B}>0$时，记为 $\boldsymbol{A}>\boldsymbol{B}$ 或 $\boldsymbol{B}<\boldsymbol{A}$.

本章的内容安排如下：在 §7.1，给出了 $\boldsymbol{A}^{+}\leqslant\boldsymbol{B}^{+}$ 的一些充要条件. §7.2和 §7.3讨论了著名的 Cauchy-Schwarz 不等

式和 Kantorovich 不等式的矩阵形式. 这两节主要总结了本书作者之一与其合作者的论文 Wang 和 Liski(1996)、Wang 和 Shao(1992)以及 Lin 和 Neudecker(1995)等的部分新结果. 在这些论文中,作者们提出了证明代数事实的一种统计学途径.

§7.1 $A^+\leqslant B^+$

假设 $A\geqslant B>0$,不难证明 $A^{-1}\leqslant B^{-1}$. 但是对 $A\geqslant B\geqslant 0$的情形,对任何形式的广义逆,相应的结论并不总是成立. 例如

$$A=\begin{pmatrix}2&&\\&1&\\&&0\end{pmatrix}\geqslant 0,\quad B=\begin{pmatrix}1&&\\&0&\\&&0\end{pmatrix}\geqslant 0,$$

但

$$A^+=\begin{pmatrix}\frac{1}{2}&&\\&1&\\&&0\end{pmatrix},\quad B^+=B=\begin{pmatrix}1&&\\&0&\\&&0\end{pmatrix},$$

显然 $A^+\leqslant B^+$不成立. 下面我们讨论 $A^+\leqslant B^+$的若干充要条件.

我们先证明几个引理.

引理7.1.1 设 $A\geqslant B\geqslant 0$,则 $\mathscr{M}(B)\subset\mathscr{M}(A)$.

证明 根据定义,$A\geqslant B\Leftrightarrow$对一切 x,$x'Ax\geqslant x'Bx$. 若 $x\in\mathscr{M}(A)^\perp$,则 $x'Ax=0$,因而有 $x'Bx=0$. 于是 $x\in\mathscr{M}(B)^\perp$. 这就证明了

$$\mathscr{M}(A)^\perp\subset\mathscr{M}(B)^\perp,$$

但是,上式等价于 $\mathscr{M}(B)\subset\mathscr{M}(A)$. 引理证毕.

引理7.1.2 设 $A\geqslant 0, B\geqslant 0$. 则 $A\geqslant B$ 当且仅当 $\mathscr{M}(B)\subset\mathscr{M}(A)$ 和 $A(A-B)A\geqslant 0$ 同时成立.

证明 结合上一引理，必要性的证明是容易的. 下面证明充分性. 因为 $\mathscr{M}(AA^{+})=\mathscr{M}(A)$，于是

$$\begin{aligned}&A(A-B)A\geqslant 0\\ &\Leftrightarrow x^{*}A(A-B)Ax\geqslant 0, \text{对一切 } x\in C^{n}\\ &\Leftrightarrow x^{*}AA^{+}(A-B)AA^{+}x\geqslant 0, \text{对一切 } x\in C^{n}.\end{aligned}$$

这就证明了

$$AA^{+}(A-B)AA^{+}\geqslant 0. \qquad (7.1.1)$$

因为 AA^{+} 是向 $\mathscr{M}(A)$ 的正交投影阵，故有 $AAA^{+}=A$. 再利用 $\mathscr{M}(B)\subset\mathscr{M}(A)$，有 $AA^{+}B=B, BAA^{+}=B$. 于是(7.1.1)等价于 $A\geqslant B$. 证毕.

引理7.1.3 设 A、B 皆为 Hermite 阵，$A\geqslant 0, B\geqslant 0$. 则 $A\geqslant B$ 当且仅当 $\mathscr{M}(B)\subset\mathscr{M}(A)$，$\lambda_1(BA^{-})\leqslant 1$，这里 $\lambda_1(BA^{-})$ 为 BA^{-} 的最大特征值，并且它与 A^{-} 的选择无关.

证明 依据引理7.1.1，问题归结为在条件 $\mathscr{M}(B)\subset\mathscr{M}(A)$ 下，证明 $A\geqslant B$ 与 $\lambda_1(BA^{-})\leqslant 1$ 等价. 再利用引理7.1.2，问题进一步化为在条件 $\mathscr{M}(B)\subset\mathscr{M}(A)$ 下证明

$$A(A-B)A\geqslant 0\Leftrightarrow\lambda_1(BA^{-})\leqslant 1. \qquad (7.1.2)$$

设 A, B 为 n 阶阵. 对 A 作满秩分解：$A=TT^{*}$，这里 T 为 $n\times r$ 矩阵，$r=r(A)$. 于是

$$A(A-B)A\geqslant 0\Leftrightarrow T^{*}(A-B)T\geqslant 0.$$

再用 $(T^{*}T)^{-1}$ 去左乘和右乘上式，并记 $U=T(T^{*}T)^{-1}$，得

$$\begin{aligned}&\Leftrightarrow I-U^{*}BU\geqslant 0\\ &\Leftrightarrow\lambda_1(U^{*}BU)\leqslant 1\\ &\Leftrightarrow\lambda_1(BUU^{*})\leqslant 1\\ &\Leftrightarrow\lambda_1(BA^{+})\leqslant 1,\end{aligned}$$

这里利用了 $A^+=T(T^*T)^{-2}T^*=UU^*$，(7.1.2)得证.

最后我们还需要证明：对一切 A^-，$\lambda_1(BA^-)=\lambda_1(BA^+)$. 事实上，因为 $\mathscr{M}(B)\subset\mathscr{M}(A)$，我们有 $\mathscr{M}(B^{\frac{1}{2}})\subset\mathscr{M}(A)$，故 $B^{\frac{1}{2}}A^-B^{\frac{1}{2}}$ 与广义逆 A^- 的选择无关. 于是

$$\lambda_1(BA^-)=\lambda_1(B^{\frac{1}{2}}A^-B^{\frac{1}{2}})=\lambda_1(B^{\frac{1}{2}}A^+B^{\frac{1}{2}})=\lambda_1(BA^-).$$

这就完成了定理的证明.

现在来证明本节的主要定理(Milliken 等，1977).

定理7.1.1 设 A,B 为同阶半正定 Hermite 阵. 则下列任两条可推出第三条.

(a) $A\geqslant B$；

(b) $B^+\geqslant A^+$；

(c) $r(A)=r(B)$.

证明 证明分如下三部分.

(1) (a)和(b)⇒(c). 应用引理7.1.1，得

$$A\geqslant B\Rightarrow\mathscr{M}(B)\subset\mathscr{M}(A),$$
$$B^+\geqslant A^+\Rightarrow\mathscr{M}(A^+)\subset\mathscr{M}(B^+).$$

但是 $\mathscr{M}(A)=\mathscr{M}(A^+)$，$\mathscr{M}(B)=\mathscr{M}(B^+)$，于是

$$\mathscr{M}(B)\subset\mathscr{M}(A)=\mathscr{M}(A^+)\subset\mathscr{M}(B^+)=\mathscr{M}(B),$$

这就证明了 $\mathscr{M}(A)=\mathscr{M}(B)$，它自然蕴含了 $r(A)=r(B)$.

(2) (a)和(c)⇒(b). 利用 $\mathscr{M}(B)=\mathscr{M}(B^+)$，$\mathscr{M}(A)=\mathscr{M}(A^+)$，从引理7.1.3得

$$\begin{aligned}A\geqslant B&\Leftrightarrow\mathscr{M}(B)\subset\mathscr{M}(A),\lambda_1(BA^+)\leqslant 1\\&\Leftrightarrow\mathscr{M}(B^+)\subset\mathscr{M}(A^+),\lambda_1(BA^+)\leqslant 1\\&\Leftrightarrow\mathscr{M}(B^+)=\mathscr{M}(A^+),\lambda_1(BA^+)\leqslant 1,\end{aligned}\qquad(7.1.3)$$

这里在最后一步利用了 $r(A^+)=r(A)=r(B)=r(B^+)$. 注意到 $\lambda_1(BA^+)=\lambda_1(A^+(B^+)^+)$，再次利用引理7.1.3，得知

(7.1.3)等价于(b).

(3) (b)和(c)⇒(a). 在第二部分的证明中,我们实际上证明了在条件(c)下:(a)⇔(b). 所以也就证明了(b)和(c)⇒(a). 定理证毕.

这个定理表明,由 $A\geqslant B$,一般并不能推出 $A^{+}\leqslant B^{+}$,因而更不能推出 $A^{(i,j,\cdots,l)}\leqslant B^{(i,j,\cdots,l)}$. 但 Wu(1980)证明了,对任意半正定 Hermite 阵 A 和 B,若 $A\geqslant B$,则一定存在 A^{-} 和 B^{-},使得 $A^{-}\leqslant B^{-}$ 成立,类似的结果对其它一些$\{i,j,\cdots,l\}$-逆也成立.

§7.2 Cauchy-Schwarz 型矩阵不等式

本节所讨论的矩阵和向量都是实的. 假设 $x,y\in R^{n}$,著名的 Cauchy-Schwarz 不等式为

$$(x'y)^{2}\leqslant(x'x)(y'y). \tag{7.2.1}$$

若 $A\in R^{n\times n}$,$A>0$,从(7.2.1)容易推出

$$(x'y)^{2}\leqslant x'Ax\cdot y'A^{-1}y. \tag{7.2.2}$$

当 $x=y$ 且 $x'x=1$时,上式可改写为

$$(x'Ax)^{-1}\leqslant x'A^{-1}x. \tag{7.2.3}$$

Marshall 和 Olkin(1990)将上式推广为如下矩阵形式:

$$(X'AX)^{-1}\leqslant X'A^{-1}X, \tag{7.2.4}$$

这里 X 为 $n\times p$ 矩阵,且满足 $X'X=I_{p}$(证明见王松桂等(1994),p.217或推论7.2.2). 近年来许多学者对(7.2.4)做了种种形式的推广. 本书作者之一与 Liski(见 Wang 和 Liski(1996)用最小二乘统一理论(见本书定理9.3.5)导出了一个颇有用的偏序关系,应用它可以很容易获得许多重要的矩阵不等式. 主要结果总结为下面的几个定理和推论.

定理7.2.1 设 $V\in R^{n\times n}, V\geqslant 0, X\in R^{n\times p}$. 则对任意一对满足

$$AX = BX \tag{7.2.5}$$

的 $k\times n$ 矩阵 A 和 B，有

$$BX(X'T^- X)^- X'B' \leqslant ATA', \tag{7.2.6}$$

其中 $T=V+XUX'$，$U\in R^{p\times p}$的对称阵，$U\geqslant 0, r(T)=r(X\vdots V)$. 进一步若 $\mathscr{M}(X)\subset\mathscr{M}(V)$，则

$$BX(X'V^- X)^- X'B' \leqslant AVA', \tag{7.2.7}$$

这里(7.2.6)和(7.2.7)的左端与所含广义逆的选择无关，且此两式等号成立当且仅当存在矩阵 H，使得

$$VA' = XH. \tag{7.2.8}$$

证明 考虑线性统计模型

$$y = X\beta + \varepsilon, E(\varepsilon) = 0, \mathrm{Cov}(\varepsilon) = V, \tag{7.2.9}$$

这里 y 为 $n\times 1$观测向量，X 和 V 如定理所设，β 为 $p\times 1$未知参数向量，ε 为 $n\times 1$随机误差向量. 记号 $E(\varepsilon)$表示 ε 的均值向量，$\mathrm{Cov}(\varepsilon)$表示 ε 的协方差矩阵.

记 $\mu=X\beta$，设 Ay 为 $B\mu$ 的无偏估计，则由 $E(Ay)=B\mu=BX\beta$，对一切 $\beta\in R^p$ 成立，可推得 A,B 满足(7.2.5). 再记

$$\tilde{\mu}= X(X'T^- X)^- X'T^- y,$$

依定理9.3.5知，$B\tilde{\mu}$为 $B\mu$ 的最佳线性无偏估计(Best Linear Unbiased Estimator，简记为 BLUE). 其协方差阵为

$$\mathrm{Cov}(B\tilde{\mu}) = B[X(X'T^- X)^- X' - XUX']B',$$

于是 $\mathrm{Cov}(B\tilde{\mu})\leqslant\mathrm{Cov}(Ay)$，即

$$B[X(X'T^- X)^- X' - XUX']B' \leqslant AVA'.$$

此式等价于(7.2.6). 若 $\mathscr{M}(X)\subset\mathscr{M}(V)$，可取 $U=0$，于是 $T=V$，这时(7.2.6)就化为(7.2.7). 再应用引理9.3.2可推出，为 $B\mu$ 的 BLUE 当且仅当存在矩阵 H，使得(7.2.8)成立，

即(7.2.6)和(7.2.7)等号成立的充要条件为(7.2.8).定理证毕.

(7.2.5)是一个非常弱的条件,我们可以找到很多的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足它,因而从(7.2.6)和(7.2.7)可以得到许多矩阵不等式,它们都可以看成(7.2.4)的推广形式.为符号简单计,在下面的推论中,我们用 $\boldsymbol{A}$ 代替了定理中的 $\boldsymbol{V}$,并且只考虑 $\mathscr{M}(\boldsymbol{X})\subset\mathscr{M}(\boldsymbol{A})$ 的情形.对一般的情况,读者不难写出相应的不等式.

推论7.2.1 设 $\boldsymbol{A}\in R^{n\times n}$, $\boldsymbol{X}\in R^{n\times p}$, $\boldsymbol{A}\geqslant 0$, $\mathscr{M}(\boldsymbol{X})\subset\mathscr{M}(\boldsymbol{A})$.

(a) 对任意的 $\boldsymbol{B}\in R^{q\times n}$ 和 $\boldsymbol{Y}\in R^{q\times k}$,有

$$\boldsymbol{Y}'\boldsymbol{B}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^-\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{B}\boldsymbol{Y}\leqslant\boldsymbol{Y}'\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}'\boldsymbol{Y},\qquad(7.2.10)$$

等号成立⇔对某个矩阵 $\boldsymbol{H}$, $\boldsymbol{A}\boldsymbol{B}'\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{H}$ 成立.

(b) $$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^-\boldsymbol{X})^-\boldsymbol{X}'\leqslant\boldsymbol{A},\qquad(7.2.11)$$

等号成立⇔$\mathscr{M}(\boldsymbol{X})=\mathscr{M}(\boldsymbol{A})$.

(c) 对任意 $\boldsymbol{X}^-$,有

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^-\boldsymbol{X})^-\boldsymbol{X}'\leqslant\boldsymbol{X}\boldsymbol{X}^-\boldsymbol{A}(\boldsymbol{X}^-)'\boldsymbol{X}',\qquad(7.2.12)$$

等号成立⇔存在矩阵 $\boldsymbol{H}$,使得 $\boldsymbol{A}(\boldsymbol{X}\boldsymbol{X}^-)'=\boldsymbol{X}\boldsymbol{H}$.

(d) 若 $\boldsymbol{X}'\boldsymbol{X}$ 可逆,则

$$(\boldsymbol{X}'\boldsymbol{A}^-\boldsymbol{X})^-\leqslant\boldsymbol{X}^-\boldsymbol{A}(\boldsymbol{X}^-)'.\qquad(7.2.13)$$

(e) 若 $\boldsymbol{W}\geqslant 0$ 且使得 $r(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})=r(\boldsymbol{X})$,则

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{A}^-\boldsymbol{X})^-\boldsymbol{X}'\leqslant\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^-\boldsymbol{X}',$$

等号成立⇔存在矩阵 $\boldsymbol{H}$,使得

$$\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^-\boldsymbol{X}'=\boldsymbol{X}\boldsymbol{H}.$$

证明 (a) 在(7.2.7)中命 $\boldsymbol{A}=\boldsymbol{B}=\boldsymbol{Y}'\boldsymbol{D}$,再将 $\boldsymbol{D}$ 和 $\boldsymbol{V}$ 分别改记为 $\boldsymbol{B}$ 和 $\boldsymbol{A}$,即得(7.2.10).相应地,(7.2.8)变为 $\boldsymbol{A}\boldsymbol{B}'\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{H}$.于是(a)得证.

(b) 在(7.2.7)中选 $A=B=I$,再将 V 改记为 A,得(7.2.11).

(c) 在(7.2.7)中,取 $A=XX^-$,$B=I$. 显然它们满足(7.2.5). 将 V 改记为 A,得(7.2.12).

(d) 若 $X'X$ 可逆,用 $(X'X)^{-1}X'$ 和 $X(X'X)^{-1}$ 分别左乘和右乘(7.2.12)的两边,便得到(7.2.13).

(e) 因为在条件 $r(X'WX)=r(X)$ 下,$X(X'WX)^-X'WX=X$,于是在(7.2.7)中取 $A=B=X(X'WX)^-X'W$,并改 V 为 A 便得到所要不等式. 推论证毕.

在(7.2.12)中,左端与所含的广义逆选择无关,而右端却含广义逆有关. 若特别取 Moore-Penrose 广义逆,便有

$$X(X'A^+X)^+X' \leqslant XX^+A(X^+)'X'.$$

推论7.2.2 设 $A>0$,$X'X=I_p$. 则

$$(X'AX)^{-1} \leqslant X'A^{-1}X. \tag{7.2.14}$$

证明 因为 $A>0$,故 $\mathscr{M}(X)\subset\mathscr{M}(A)$ 自然成立. 用 X' 和 X 分别左乘和右乘(7.2.11)两边,得

$$(X'A^{-1}X)^{-1} \leqslant X'AX,$$

它等价于(7.2.14). 证毕.

(7.2.14)就是 Marshall 和 Olkin(1990)所证明的不等式.

在上面的定理和推论中,我们分别要求 $\mathscr{M}(X)\subset\mathscr{M}(V)$ 或 $\mathscr{M}(X)\subset\mathscr{M}(A)$ 成立. 下面放弃这个假设,建立另外一类等式.

定理7.2.2 设 $V\in R^{n\times n}$,$V\geqslant 0$,$X\in R^{n\times p}$. 记 P_V 为向 $\mathscr{M}(V)$ 的正交投影阵. 则对任意一对满足

$$AP_VX = BP_VX \tag{7.2.15}$$

的 $k\times n$ 矩阵 A 和 B,有

$$BP_VX(X'V^+X)^+X'P_VB' \leqslant AVA', \qquad (7.2.16)$$

等号成立$\Leftrightarrow$存在矩阵 H,使得

$$VA' \doteq P_VXH. \qquad (7.2.17)$$

证明 因为 $\mathscr{M}(P_VX)\subset\mathscr{M}(V)$,在定理7.2.1中用 P_VX 代替 X,并利用

$$P_VV^-P_V = V^+,$$

(7.2.5)、(7.2.7)和(7.2.8)分别化为(7.2.15)、(7.2.16)和(7.2.17),定理证毕.

在(7.2.16)中分别选择 $A=B=Y'$ 和 $A=B=Y'V$,并改 V 为 A,便得到下面的推论.

推论7.2.3 设 $A\in R^{n\times n}$,$A\geqslant 0$,$X\in R^{n\times p}$,$Y\in R^{n\times q}$. 则

(a) $Y'P_AX(X'A^+X)^+X'P_AY\leqslant Y'AY$,

等号成立当且仅当,存在矩阵 H 使得 $AY'=P_AXH$.

(b) $Y'AX(X'A^+X)^+X'P_AY\leqslant Y'A^3Y$,

等号成立当且仅当,存在矩阵 H 使得 $A^2Y=P_AXH$.

下面一个推论包含了半正定阵的幂,为此,我们先引进如下定义.

设 $A\in R^{n\times n}$,$A\geqslant 0$. 则存在 $n\times n$ 正交阵 Q,使得

$$A = Q\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q',$$

这里 $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_r)$,$\lambda_i>0$,$i=1,\cdots,r$,$r=r(A)$. 定义

$$A^s = Q\begin{pmatrix} \Lambda^s & 0 \\ 0 & 0 \end{pmatrix}Q',$$

这里 s 为任一实数. 根据这个定义,容易验证 $A^0=P_A$,且当 $r(A)=n$ 时,$A^0=I_n$.

在(7.2.16)和(7.2.17)中,令 $X=V^{\frac{s+1}{2}}Z$,$A=B=Y$,然后将 V 和 Z 分别换成 A 和 X,便得到如下结论.

推论7.2.3 设$A\in R^{n\times n}$，$A\geqslant 0$，$X\in R^{n\times p}$，$Y\in R^{n\times q}$. 则

$$Y'A^{\frac{k+1}{2}}X(X'A^kX)^+X'A^{\frac{k+1}{2}}Y\leqslant Y'AY,$$

等号成立当且仅当存在矩阵H使得

$$AY=A^{\frac{k+1}{2}}XH,$$

这里k为任一实数.

对矩阵A、B其它的选择，还可导出一些有意义的矩阵不等式，详细讨论见 Wang 和 Liski(1996).

§7.3 Kantorovich 型矩阵不等式

设$A\in R^{n\times n}$，$A>0$，x为$n\times 1$实向量，满足$x'x=1$. 记$\lambda_1\geqslant\cdots\geqslant\lambda_n$为$A$的特征值. 则

$$x'Axx'A^{-1}x\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}. \tag{7.3.1}$$

这就是著名的 Kantorovich 不等式. 关于(7.3.1)，近年来有种种形式的推广. 一类推广是将向量x换成矩阵X或Y，然后研究其行列式的上界，Bloomfiled 和 Watson(1975)与 Khatri 和 Rao(1981)的工作就属这一类. 另一类是将向量x换成矩阵X之后，考虑形如

$$X'A^{-1}X\leqslant\frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X'AX)^{-1} \tag{7.3.2}$$

的偏序以及其它变形，这里假定$X'X=I_p$. (7.3.2)是由 Marshall 和 Olkin(1990)证明的. 本节将讨论(7.3.2)种种形式推广的某些重要新结果. 主要材料取自本书作者之一与邵军(见 Wang 和 Shao(1992))以及 Liu 和 Neudecker(1995)的论文. 和上节相类似，我们将利用统计定理证明代数结论.

我们先证明几个引理.

引理7.3.1 (Kantorovich 不等式) 设 $A \in R^{n \times n}, A>0$, 则对任意 $n \times 1$ 非零向量 x, 有

$$\frac{x'Axx'A^{-1}x}{(x'x)^2} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}, \qquad (7.3.3)$$

这里 λ_1 和 λ_n 为 A 的最大和最小特征值.

证明 设 Q 为 $n \times n$ 正交阵, 使 $A=Q\Lambda Q'$, 这里 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ 为 A 的特征值. 记 $u=Q'x$, 则(7.3.3)等价于

$$u'\Lambda u \cdot u'\Lambda^{-1}u \leqslant \frac{\lambda_1+\lambda_n}{2} \cdot \frac{\lambda_1^{-1}+\lambda_n^{-1}}{2}u'u$$

$$\Leftrightarrow u'\left(\frac{2}{\lambda_1+\lambda_n}\Lambda\right)u \cdot u\left(\frac{2}{\lambda_1^{-1}+\lambda_n^{-1}}\Lambda^{-1}\right)u \leqslant u'u.$$

利用几何平均不超过算术平均, 可推出上式成立的一个充分条件为

$$u'\left(\frac{\Lambda}{\lambda_1+\lambda_n}+\frac{\Lambda^{-1}}{\lambda_1^{-1}+\lambda_n^{-1}}\right)u \leqslant u'u,$$

但此式等价于

$$\frac{\lambda_i}{\lambda_1+\lambda_n}+\frac{\lambda_i^{-1}}{\lambda_1^{-1}+\lambda_n^{-1}} \leqslant 1, \quad i=1, \cdots, n$$

$$\Leftrightarrow (\lambda_i-\lambda_1)(\lambda_i-\lambda_n) \leqslant 0, \quad i=1, \cdots, n.$$

引理证毕.

为了引进下一个引理, 我们考虑如下线性统计模型:

$$y=X\beta+\varepsilon, E(\varepsilon)=0, \mathrm{Cov}(\varepsilon)=V, \qquad (7.3.4)$$

这里 y 为 $n \times 1$ 观测向量, X 为 $n \times p$ 矩阵, $r(X)=r \leqslant p$, β 为 $p \times 1$ 未知参数向量, ε 为 $n \times 1$ 随机误差, $E(\varepsilon)$ 和 $\mathrm{Cov}(\varepsilon)$ 分别表示 ε 的均值向量和协方差阵. 假定 $V>0$.

若 $c \in \mathscr{M}(X')$, 我们称 $c'\beta$ 为可估函数, 因为此时存在向量 a, 使得 $a'y$ 为 $c'\beta$ 的无偏估计. $c'\beta$ 的最小二乘估计(Least

Squares Estimator,简记为 LSE)及其方差分别为

$$c'\hat{\beta} = c'(X'X)^{+}X'y$$

和

$$\mathrm{Var}(c'\hat{\beta}) = c'(X'X)^{+}X'VX(X'X)^{+}c.$$

而 $c'\beta$ 的 BLUE 及其方差分别为

$$c'\tilde{\beta} = c'(X'V^{-1}X)^{+}X'V^{-1}y$$

和

$$\mathrm{Var}(c'\tilde{\beta}) = c'(X'V^{-1}X)^{+}c.$$

$c'\hat{\beta}$关于 $c'\tilde{\beta}$的相对效率定义为

$$\mathrm{RE}(c'\hat{\beta}) = \frac{\mathrm{Var}(c'\tilde{\beta})}{\mathrm{Var}(c'\hat{\beta})},$$

它度量了 LSE 的优劣程度.下面的引理给出了 $\mathrm{RE}(c\hat{\beta})$的一个下界.

引理7.3.2 在线性统计模型(7.3.4)中,设 $V>0$,$\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 V 的特征值,则

$$\mathrm{RE}(c'\hat{\beta}) \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}. \tag{7.3.5}$$

证明 设

$$X = P\Delta Q' \tag{7.3.6}$$

为 X 的奇异值分解,这里 $P\in R^{n\times r}$,$Q\in R^{p\times r}$且 $P'P=I_r$,$Q'Q=I_r$,

$$\Delta = \mathrm{diag}(\delta_1,\cdots,\delta_r),\delta_i>0,i=1,\cdots,r,$$

$r=r(A)$.因为 $c'\beta$ 可估,所以存在 α,使得 $c=X'\alpha$.于是

$$\mathrm{RE}(c'\hat{\beta}) = \frac{\alpha'X(X'V^{-1}X)^{+}X'\alpha}{\alpha'P_XVP_X\alpha}, \tag{7.3.7}$$

其中 $P_X=X(X'X)^{+}X'$.将(7.3.6)代入(7.3.7)并令 $u=P'\alpha$,

得

$$\mathrm{RE}(c'\hat{\beta}) = \frac{u'(P'V^{-1}P)^{-1}u}{u'PVPu}. \qquad (7.3.8)$$

在 Cauchy-Schwarz 不等式(7.2.2)中，取 $A=P'V^{-1}P$，我们有

$$(u'u)^2 \leqslant u'P'V^{-1}Pu \cdot u'(P'V^{-1}P)^{-1}u,$$

等价地

$$u'(P'V^{-1}P)^{-1}u \geqslant \frac{(u'u)^2}{u'P'V^{-1}Pu}.$$

代入(7.3.8)，并令 $w=Pu$，利用 $w'w=u'u$，(7.3.8)变为

$$\mathrm{RE}(c'\hat{\beta}) \geqslant \frac{(w'w)^2}{w'V^{-1}w \cdot w'Vw} \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2},$$

最后一步利用了引理7.3.1. 证毕.

假设 A 和 B 皆为 n 阶正定阵. 在线性模型(7.3.4)中，用 A^{-1}代替 V，并用 $B^{1/2}$左乘，得到新模型

$$B^{1/2}y = B^{1/2}X\beta + \varepsilon, E(\varepsilon) = 0, \mathrm{Cov}(\varepsilon) = B^{1/2}A^{-1}B^{1/2}$$

记 $\mu=B^{1/2}X\beta$，则 μ 的 LSE 和它的协方差阵分别为

$$\hat{\mu} = B^{1/2}X(X'BX)^{+}X'By$$

和

$$\mathrm{Cov}(\hat{\mu}) = B^{1/2}X(X'BX)^{+}X'BA^{-1}BX(X'BX)^{+}X'B^{1/2}. \qquad (7.3.9)$$

而 μ 的 BLUE 和它的协方差阵分别为

$$\tilde{\mu} = B^{1/2}X(X'AX)^{+}X'Ay$$

和

$$\mathrm{Cov}(\tilde{\mu}) = B^{1/2}X(X'AX)^{+}X'B^{1/2}. \qquad (7.3.10)$$

利用引理7.3.2，Wang 和 Shao(1992)证明了如下定理.

定理7.3.1 设 $X\in R^{n\times p}$，A 和 B 皆为 n 阶正定阵，$\tau_1\geqslant\cdots$

$\geqslant \tau_n$ 为 $B^{1/2}A^{-1}B^{1/2}$的特征值. 则

$$X'BA^{-1}BX \leqslant \frac{(\tau_1+\tau_n)^2}{4\tau_1\tau_n}X'BX(X'AX)^+ X'BX. \tag{7.3.11}$$

证明 设 c 为任一 $n\times 1$实向量. 利用引理7.3.2我们有

$$\mathrm{RE}(c'\hat{\mu}) \geqslant \frac{4\tau_1\tau_n}{(\tau_1+\tau_n)^2},$$

等价地,

$$\mathrm{Var}(c'\hat{\mu}) \leqslant \frac{(\tau_1+\tau_n)^2}{4\tau_1\tau_n}\mathrm{Var}(c'\tilde{\mu}).$$

由 c 的任意性,从上式可得

$$\mathrm{Cov}(\hat{\mu}) \leqslant \frac{(\tau_1+\tau_n)^2}{4\tau_1\tau_n}\mathrm{Cov}(\tilde{\mu}).$$

将(7.3.9)、(7.3.10)代入上式后,两边分别左乘 $X'B^{1/2}$,右乘 $B^{1/2}X$,并利用 $X'BX(X'BX)^+XB=XB$,便得到(7.3.11). 定理证毕.

在定理中若取 $B=I$,则得到如下推论.

推论7.3.1 设 $A\in R^{n\times n}$,$A>0$,$\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 A 的特征值. $X\in R^{n\times p}$,$r(X)\leqslant p$. 则

$$X'A^{-1}X \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}X'X(X'AX)^+ X'X. \tag{7.3.12}$$

若 $X'X$ 可逆,则

$$(X'X)^{-1}X'A^{-1}X(X'X)^{-1} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X'AX)^{-1}. \tag{7.3.13}$$

特别当 $X'X=I_p$ 时,上式变为

$$X'A^{-1}X \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X'AX)^{-1}. \tag{7.3.14}$$

(7.3.12)的右端的 Moore-Penrose 广义逆可以换成{1}-逆，这就是说，(7.3.12)的右端与所含广义逆的选择无关. 关于这个简单事实的证明，留给读者做练习.

在定理7.3.1中，令 $\boldsymbol{A}=\boldsymbol{I}$ 并将 $\boldsymbol{B}$ 换成 $\boldsymbol{A}$，我们得到

推论7.3.2 在推论7.3.1的条件下，有

$$\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\boldsymbol{X}'\boldsymbol{A}\boldsymbol{P}_X\boldsymbol{A}\boldsymbol{X},$$

这里 $\boldsymbol{P}_X=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. 若 $\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{I}_p$，则

$$\boldsymbol{X}'\boldsymbol{A}^2\boldsymbol{X} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X})^2.$$

推论7.3.3 设 $\boldsymbol{A}\in\boldsymbol{R}^{n\times n}$，$\boldsymbol{A}>0$. 将 $\boldsymbol{A}$ 与 $\boldsymbol{A}^{-1}$分块为

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix},\boldsymbol{A}^{-1}=\begin{pmatrix}\boldsymbol{A}^{11} & \boldsymbol{A}^{12}\\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22}\end{pmatrix},$$

其中 $\boldsymbol{A}_{11}$和 $\boldsymbol{A}^{11}$均为 $m\times m$ 矩阵. 则

$$\boldsymbol{A}^{11} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}\boldsymbol{A}_{11}^{-1}. \qquad (7.3.15)$$

证明 在(7.3.14)中，取

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{I}_m\\ \boldsymbol{0}\end{pmatrix},$$

便得到(7.3.15). 证毕.

Liu 和 Neudecker(1995)应用 Wang 和 Shaò(1992)的方法，将上面的结果推广到 $\boldsymbol{A}\geqslant 0$的情形. 他们的主要结果可归纳为下面的定理.

定理7.3.2 设 $\boldsymbol{A}\in\boldsymbol{R}^{n\times n}$，$\boldsymbol{X}\in\boldsymbol{R}^{n\times p}$.

(a) 若 $\boldsymbol{A}>0$，$\lambda_1\geqslant\cdots\geqslant\lambda_n$ 为 $\boldsymbol{A}$ 的特征值. 则

$$\boldsymbol{X}^+\boldsymbol{A}\boldsymbol{X}^{+\prime} \leqslant \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(\boldsymbol{X}'\boldsymbol{A}^{-1}\boldsymbol{X})^-. \qquad (7.3.16)$$

(b) 若 $A \geqslant 0$，$\lambda_1 \geqslant \cdots \geqslant \lambda_r$ 为 A 的非零特征值，$\mathscr{M}(X) \subset \mathscr{M}(A)$. 则

$$X^+ AX^{+\prime} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(X'A^+ X)^+. \qquad (7.3.17)$$

证明 (a) 在线性模型(7.3.4)中，将 V 改为 A，利用引理7.3.2得

$$c'(X'X)^+ X'AX(X'X)^+ c \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} c'(X'A^{-1}X)^+ c.$$

因为存在 α，使得 $c = X'\alpha$，故上式可变形为

$$\alpha' P_X A P_X \alpha \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \alpha' X(X'A^{-1}X)^+ X'\alpha.$$

事实上，上面的不等式对一切 $\alpha \in R^n$ 都成立，再利用 $P_X = XX^+$，得

$$XX^+ AXX^+ \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} X(X'A^{-1}X)^+ X'.$$

因为 $XX^+ = (XX^+)' = (X^+)'X'$，上式可进一步改写为

$$XX^+ A(X^+)'X' \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} X(X'A^{-1}X)^+ X'.$$

此式等价于(7.3.16).

(b) 若 $A \geqslant 0$，$r(A) = r$. 则存在 $Q \in R^{n \times r}$，$Q'Q = I_r$，使得

$$A = Q\Lambda Q',$$
$$A^- = Q\Lambda^{-1}Q',$$

其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_r)$，定义 $Y = Q'X \in R^{r \times p}$. 因为 $\Lambda > 0$，在(7.3.16)中用 Λ 代替 A，用 Y 代替 X，得

$$Y^+ \Lambda Y^{+\prime} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(Y'\Lambda^{-1}Y)^+. \qquad (7.3.18)$$

利用定理4.2.2(d)易证 $X^+ = Y^+Q'$，所以

$$Y^+ \Lambda Y^{+\prime} = Y^+ Q'AQY^{+\prime} = X^+ AX^{+\prime},$$

$$Y'\Lambda^{-1}Y = X'Q\Lambda^{-1}QX = X'A^{+}X.$$

将此两式代入(7.3.18)，即得(7.3.17).定理证毕.

关于这一方面的更多结果，读者可阅读 Wang 和 Shao (1992)以及 Liu 和 Neudecker(1995).

第八章　广义逆的计算

关于普通逆矩阵的计算，文献中已经有了许多行之有效的方法．广义逆矩阵作为普通逆矩阵概念的推广，当我们要计算它时，很自然希望把它归结为通常逆矩阵的计算，这是本章要讨论的许多计算方法的核心．另一方面，矩阵维数直接涉及到算法的计算量的大小．因此，降低所要计算的逆矩阵的维数，是一个很重要的问题．在这里，分块矩阵是一个有力的工具．

本章主要讨论 A^+ 的计算．因为 A^+ 也是其它 $\{i,j,\cdots,l\}$-逆，所以求得 A^+，也就得到了一个 $A^{(i,j,\cdots,l)}$．我们侧重于介绍计算公式，其中一些，把计算 A^- 或其它广义逆归结为普通逆的计算，另一些给出了迭代公式．有了这些公式，具体计算就可以利用计算方法中已有结果(参阅王国荣(1994))．

§8.1　基于满秩分解的方法

矩阵的满秩分解在广义逆的计算中，占有特殊重要的地位．其主要原因是利用它可以很容易把广义逆计算归结为普通逆矩阵的计算．

我们用 $C_r^{m\times n}$ 表示秩为 r 的 $m\times n$ 阵的全体．设 $A\in C_r^{m\times n}$，则存在 $F\in C_r^{m\times r}$ 和 $G\in C_r^{r\times n}$，使得

$$A=FG,\tag{8.1.1}$$

这就是所谓的满秩分解(见定理 2.4.5)．根据定理 4.1.3 及推论，我们知道

$$A^+ = G^*(F^*AG^*)^{-1}F^* \tag{8.1.2}$$

$$= G^*(GG^*)^{-1}(F^*F)^{-1}F^*. \tag{8.1.3}$$

特别，(a) 当 $A\in C_r^{m\times r}$ 时

$$A^+ = (A^*A)^{-1}A^*, \tag{8.1.4}$$

(b) 当 $A\in C_r^{r\times n}$ 时

$$A^+ = A^*(AA^*)^{-1}. \tag{8.1.5}$$

这里给出了利用满秩分解求 A^+ 的计算公式，自然，它也提供了计算出一个 $A^{(1)}, A^{(1,2)}, A^{(1,3)}, A^{(1,2,4)}, A^{(1,2,3)}$ 等 $\{i,j,\cdots,l\}$-广义逆的方法．

注 对于求一个矩阵的满秩分解，文献中有许多方法，例如化阶梯形法、高斯消去法、Householder 变换法等．这些可以在一般计算方法的书中找到．

§8.2 基于分块矩阵的方法

设 $A\in C_r^{m\times n}$，因为 $r(A)=r$，于是至少存在一个 r 阶子阵是可逆的．用互换 A 的行与行和列与列，可以把这个可逆子阵移到 A 的左上角，即存在置换方阵 P 和 Q，使得

$$PAQ = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{8.2.1}$$

其中 $A_{11}\in C_r^{r\times r}$．这时 A_{22} 与 A_{11}，A_{12} 和 A_{21} 有如下关系：

引理 8.2.1 $A_{22}=A_{21}A_{11}^{-1}A_{12}$.

证明 记

$$H_1 = \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix},$$

$$H_2 = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix},$$

则
$$H_1PAQH_2=\begin{pmatrix}A_{11} & 0\\ 0 & A_{22.1}\end{pmatrix},$$
这里 $A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$. 因为 H_1 和 H_2 皆为可逆阵，于是
$$\begin{aligned} r(H_1PAQH_2) &= r(A_{11})+r(A_{22.1})=r(A)\\ &= r(A_{11})=r. \end{aligned}$$
因而 $r(A_{22.1})=0$，即 $A_{22}=A_{21}A_{11}^{-1}A_{12}$. 引理证毕.

因为 $r(A)=r(A_{11})=r$，于是存在 $T\in C^{r\times(n-r)}$ 和 $S\in C^{(m-r)\times r}$，使得
$$\begin{pmatrix}A_{12}\\ A_{22}\end{pmatrix}=\begin{pmatrix}A_{11}\\ A_{21}\end{pmatrix}T,\ (A_{21}\ \vdots\ A_{22})=S(A_{11}\ \vdots\ A_{12}). \quad (8.2.2)$$
利用引理 8.2.1，得
$$A_{22}=A_{21}A_{11}^{-1}A_{12}=SA_{11}A_{11}^{-1}A_{11}T=SA_{11}T,$$
所以 PAQ 可改写为
$$PAQ=\begin{pmatrix}I\\ S\end{pmatrix}A_{11}(I\ \vdots\ T). \quad (8.2.3)$$
注意到 P 和 Q 为置换阵，有 $P'=P^{-1}$，$Q'=Q^{-1}$. 因而
$$A=P'\begin{pmatrix}I\\ S\end{pmatrix}A_{11}(I\ \vdots\ T)Q'. \quad (8.2.4)$$
利用这个分解式，我们可以证明如下定理：

定理 8.2.1 设 $A\in C_r^{m\times n}$，且可表为(8.2.4)，其中 $A_{11}\in C_r^{r\times r}$，$P$，$Q$ 为置换阵. 则

(a) A 的一个{1,2}-逆为
$$A^{(1,2)}=Q\begin{pmatrix}A_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}P.$$

(b) A 的一个{1,2,3}-逆为
$$A^{(1,2,3)}=Q\begin{pmatrix}A_{11}^{-1}\\ 0\end{pmatrix}(I+S^*S)^{-1}(I+S^*)P.$$

(c) A 的一个{1,2,4}-逆为

$$A^{(1,2,4)} = Q\begin{pmatrix} I_r \\ T^* \end{pmatrix}(I_r + TT^*)^{-1}(A_{11}^{-1} \vdots 0)P.$$

(d) $$A^+ = Q\begin{pmatrix} I_r \\ T^* \end{pmatrix}(I_r + TT^*)^{-1}A_{11}^{-1}(I_r - S^*S)^{-1} \cdot (I_r \vdots S^*)P.$$

证明 记

$$F = P'\begin{pmatrix} I_r \\ S \end{pmatrix}A_{11}, \quad G = (I_r \vdots T)Q',$$

则(8.2.4)变为

$$A = FG.$$

显然 $F \in C_r^{m\times r}$, $G \in C_r^{r\times n}$. 故上式是 A 的满秩分解. 利用(8.1.3),立即得到(d). 前面三条可以用定义去直接验证. 定理证毕.

(d)给出的计算 A^+ 的公式,称为 Noble 公式(见 Noble, 1966).

定理 8.2.2 设 $A \in C_r^{m\times n}$,且分块为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

其中 $r(A_{11}) = r(A) = r$,则

$$A^+ = (A_{11} \vdots A_{12})M^*\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}^*, \qquad (8.2.5)$$

这里

$$M = \left[(A_{11} \vdots A_{12})A^*\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}\right]^{-1}.$$

证明 在条件 $r(A) = r(A_{11})$ 下,利用(8.2.2)将 A 改写为

$$
\begin{aligned}
A=\begin{pmatrix} A_{11} & A_{12} \\ SA_{11} & SA_{12} \end{pmatrix} &= \begin{pmatrix} I_r \\ S \end{pmatrix} (A_{11} \;\vdots\; A_{12}) \\
&= FG,
\end{aligned}
$$

这里

$$
F=\begin{pmatrix} I_r \\ S \end{pmatrix}, \quad G=(A_{11} \;\vdots\; A_{12}).
$$

应用(8.1.2)便得到(8.2.5). 证毕.

本节的两个定理都是利用分块矩阵,把求一个大矩阵的广义逆的问题归结为求可逆子矩阵的通常逆的问题. 这样,既降低了所处理的矩阵的维数,又把求广义逆问题化为求通常逆的问题.

§8.3 基于镶边矩阵的方法

在上一节,我们讨论了基于分块矩阵求广义逆的方法. 这个方法的核心是用分块的方法把一个大矩阵剖分为小的含有可逆子矩阵的分块矩阵. 与此相反,本节是用镶边的方法把一个小的矩阵扩充为大的可逆阵,利用求大矩阵的逆矩阵,来获得小矩阵的 Moore-Penrose 广义逆矩阵.

设 $A\in C_r^{m\times n}$,考虑镶边矩阵

$$
M=\begin{pmatrix} A & X \\ Y^* & 0 \end{pmatrix}. \tag{8.3.1}
$$

对 X 和 Y 加上适当条件,可使 M 为可逆阵,利用 M^{-1}来求 A^+.

定理 8.3.1 设 $A\in C_r^{m\times n}$,$X\in C_{m-r}^{m\times(m-r)}$,$Y\in C_{n-r}^{n\times(n-r)}$,且满足

(a) $\mathscr{M}(X)=\mathscr{N}(A^*)$;

(b) $$\mathscr{M}(Y)=\mathscr{N}(A),$$

则由(8.3.1)定义的矩阵 M 是可逆的，且

$$M^{-1}=\begin{pmatrix} A^+ & Y^{*+} \\ X^+ & 0 \end{pmatrix}.$$

证明 我们要证明

$$\begin{pmatrix} AA^+ + XX^+ & AY^{*+} \\ Y^*A^+ & Y^*Y^{*+} \end{pmatrix}=I_{m+n-r}. \qquad (8.3.2)$$

由假设条件(a)及定理4.2.4，得

$$\mathscr{M}(X)=\mathscr{N}(A^*)=\mathscr{M}(A)^{\perp}.$$

因为

$$XX^+=P_{\mathscr{M}(XX^+)}=P_{\mathscr{M}(X)},$$

$$AA^+=P_{\mathscr{M}(AA^+)}=P_{\mathscr{M}(A)},$$

于是

$$AA^+ + XX^+=P_{\mathscr{M}(A)}+P_{\mathscr{M}(X)}=P_{\mathscr{M}(A)}+P_{\mathscr{M}(A)^{\perp}}=I_m. \qquad (8.3.3)$$

利用 A^+A 是Hermite阵及假设条件(b)，得

$$Y^*A^+=Y^*A^+AA^+=Y^*(A^+A)^*A^+=Y^*A^*A^{+*}A^+$$

$$=(AY)^*A^{+*}A^+=0. \qquad (8.3.4)$$

类似地，利用 YY^+ 是Hermite阵及假设条件(b)，有

$$AY^{*+}=AY^{+*}=A(Y^+YY^+)^*=A(Y^+(YY^+)^*)^*$$

$$=A(Y^+Y^{+*}Y^*)^*=AY(Y^+Y^{+*})^*=0. \qquad (8.3.5)$$

最后，由假设知，Y^* 是行满秩的，利用定理3.2.2(c)得

$$Y^*Y^{*+}=I_{n-r}. \qquad (8.3.6)$$

综合(8.3.3)—(8.3.6)就完成了(8.3.2)的证明．定理证毕.

推论8.3.1 设 $A\in C_r^{m\times n}$，$X\in C^{m\times(m-r)}$，$Y\in C^{n\times(n-r)}$，且满足

(a) $$A^* X = 0,\quad X^* X = I_{m-r};$$

(b) $$AY = 0,\quad Y^* Y = I_{n-r}.$$

则

$$M = \begin{pmatrix} A & X \\ Y^* & 0 \end{pmatrix}$$

为可逆阵，且

$$M^{-1} = \begin{pmatrix} A^+ & Y \\ X^* & 0 \end{pmatrix}.$$

证明 由(a)可知 $X \in C_{m-r}^{m\times(m-r)}$，$\mathscr{M}(X) = \mathscr{N}(A^*)$. 再由(b)推出 $Y \in C_{n-r}^{n\times(n-r)}$ 且 $\mathscr{M}(Y) = \mathscr{N}(A)$. 因此定理 8.3.1 的条件成立，故

$$M^{-1} = \begin{pmatrix} A^+ & Y^{*+} \\ X^+ & 0 \end{pmatrix}.$$

问题归结为证明：

$$Y = Y^{*+},\quad X^+ = X^*. \tag{8.3.7}$$

但是，从 X 和 Y 为列满秩及定理 4.2.2(d) 可立即推出(8.3.7). 推论证毕.

定理 8.3.1 和推论 8.3.1 表明，求 A^+ 的问题可以归结为求镶边矩阵 M 的普通逆矩阵. 在定理中要求选择 $X \in C_{m-r}^{m\times(m-r)}$，$Y \in C_{n-r}^{n\times(n-r)}$ 且满足

$$A^* X = 0,\quad AY = 0.$$

这等价于 X 的 $m-r$ 个列向量为线性方程组 $A^* x = 0$ 的 $m-r$ 个线性无关的解，而 Y 的 $n-r$ 个列向量为线性方程组 $Ay = 0$ 的 $n-r$ 个线性无关的解. 而在推论中，还进一步要求将这些解做标准正交化，即要求 $X^* X = I_{m-r}$，$Y^* Y = I_{n-r}$，这样一来，

我们把求 A^+ 的问题化为求两个线性方程组 $A^*x=0$ 和 $Ay=0$ 的基础解系以及镶边矩阵 M 的普通逆矩阵．

§8.4 迭代方法

设 $A\in C^{m\times n}$，我们知道，向 $\mathscr{M}(A)$ 的正交投影阵为

$$P_A = AA^-.$$

因此，若用序列 $X_k\in C^{n\times m}, k=0,1,2,\cdots$ 作为 A^+ 的近似，其近似程度可用残差

$$R_k = P_A - AX_k,\quad k = 0,1,2,\cdots \tag{8.4.1}$$

来度量．显然，当 $X_k\to A^+$ 时，$R_k\to 0$. 若选定 X_0，我们提出迭代公式为

$$X_{k+1} = X_k + X_0R_k,\quad k = 0,1,2,\cdots \tag{8.4.2}$$

在这个迭代公式中，R_k 包含了 P_A，因而也就包含了 A^+. 为了克服这一缺陷，我们取 X_0 满足

$$X_0 = X_0P_A, \tag{8.4.3}$$

这是很容易做到的．只要取 X_0 满足 $\mathscr{M}(X_0^*)\subset\mathscr{M}(A)$ 即可．于是

$$\begin{aligned} X_0R_k &= X_0(P_A - AX_k) \\ &= X_0(I - AX_k) \end{aligned} \tag{8.4.4}$$

代入(8.4.2)，得到迭代公式

$$X_{k+1} = X_k + X_0(I - AX_k), k = 0,1,2,\cdots \tag{8.4.5}$$

我们将证明，适当选取初始近似阵 X_0，(8.4.4)定义的序列 $X_k\to A^+$. 先证明几个引理．

设 $A\in C^{n\times n}$，$\lambda_i(A)(i=1,\cdots,n)$ 为 A 的特征值，记

$$\rho(A) = \max_{1\leqslant i\leqslant n}\{|\lambda_i(A)|\},$$

称 $\rho(A)$ 为 A 的谱半径．容易证明，对任一相容范数 $\|A\|$，

$$\rho(A) \leqslant \|A\|,$$

且
$$\rho(A^k) = \rho^k(A), \quad k = 0,1,2,\cdots$$

特别当 A 为 Hermite 阵时

$$\|A\|_2 = \rho^{1/2}(AA^*) = \rho(A).$$

引理 8.4.1 设 $A \in C^{n\times n}$，则下列三条等价：

(a) $\rho(A)<1$；

(b) $A^k \to 0$，当 $k \to \infty$；

(c) $I + A + \cdots + A^k + \cdots$ 收敛，其和为 $(I-A)^{-1}$.

证明 设 A 的 Jordan 标准形为 J，则

$$A = PJP^{-1},$$

其中
$$J = \mathrm{diag}(J_{n_1}(\lambda_1), J_{n_2}(\lambda_2), \cdots, J_{n_s}(\lambda_s)),$$

$\lambda_1,\cdots,\lambda_s$ 为 A 的全部特征值(参见定理 2.2.2)．于是

$$A^k = PJ^kP^{-1},$$

$$J^k = \mathrm{diag}\left(J_{n_1}^k(\lambda_1), J_{n_2}^k(\lambda_2), \cdots, J_{n_s}^k(\lambda_s)\right),$$

显然，
$$A^k \to 0 \Leftrightarrow J^k \to 0 \Leftrightarrow J_{n_i}^k(\lambda_i) \to 0, i = 1,\cdots,s,$$

这里

$$J_{n_i}^k(\lambda_i) = \begin{pmatrix} g_k(\lambda_i) & g'_k(\lambda_i) & \cdots & \dfrac{g_k^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ & g_k(\lambda_i) & \cdots & \dfrac{g_k^{(n_i-1)}(\lambda_i)}{(n_i-2)!} \\ & & \ddots & \vdots \\ & & & g_k(\lambda_i) \end{pmatrix},$$

其中 $g_k(\lambda)=\lambda^k$，$g_k^{(t)}(\lambda)$ 表示 $g_k(\lambda)$ 的 t 阶导数．

(a)$\Rightarrow$(b)．若 $\rho(A)<1$，必有 $|\lambda_i|<1$，$i=1,\cdots,s$，于是当 $k \to \infty$ 时，

$$g_k^{(l)}(\lambda_i) \to 0,$$

$$l = 0,1,2,\cdots,n_{i-1}, \qquad i = 1,\cdots,s.$$

故当 $k \to \infty$ 时，$J_{n_i}^k(\lambda_i) \to 0$，$\quad i = 1,\cdots,s$.

(b)⇒(a). 用反证法，若 $\rho(A)=1$，则必存在一个特征值，其模为 1，设为 λ_1，则

$$g_k^{(l)}(\lambda_1) \text{ 不趋于零}, \quad l = 0,1,2,\cdots,n_{1-1}.$$

于是 $J_{n_1}^{(k)}(\lambda_1)$也不趋于零．

(b)⇒(c). 若 $A^k \to 0$，由已证部分知 $\rho(A)<1$，所以，A 的所有特征值的模都小于 1，于是 $I-A$ 的特征值均不会等于零，因而 $I-A$ 可逆．但

$$(I + A + A^2 + \cdots + A^k)(I - A) = I - A^{k+1},$$

于是

$$I + A + A^2 + \cdots + A^k = (I - A)^{-1} - A^{k+1}(I - A)^{-1}.$$

当 $k\to\infty$时，上式右端第二项趋于零，故

$$I + A + \cdots + A^k + \cdots \to (I - A)^{-1}.$$

(c)⇒(b)　若矩阵级数

$$I + A + \cdots + A^k + \cdots$$

收敛，记其和为 S，则它的每一个元素都收敛．记 $A^k=(a_{ij}^{(k)})$，则 S 的(i,j)元为

$$S_{ij} = \delta_{ij} + a_{ij} + a_{ij}^{(2)} + \cdots + a_{ij}^{(k)} + \cdots$$

其中，当 $i\neq j$ 时 $\delta_{ij}=1$，不然为 0. 这是一个数项级数，它收敛的必要条件为当 $k\to\infty$时 $a_{ij}^{(k)}\to 0$，即 $A^k\to 0$. 引理证毕．

设 $A,B\in C^{m\times n}$定义

$$\mathscr{M}(A,B)= \{X = AYB, Y \in C^{n\times m}\},$$

$$\mathscr{N}(A,B)= \{X \in C^{n\times m}, \quad AXB = 0\},$$

分别称为(A,B)的值空间和零空间．

引理 8.4.2 设 $A\in C^{m\times n}$，则

$$X = A^+ \Leftrightarrow X \in \mathscr{M}(A^*, A^*), X \in A\{1\}.$$

证明 因为 $X\in A\{1\}\Leftrightarrow X$ 是矩阵方程 $AXA=A$ 的解，依定理 3.3.1，通解可表为

$$X = A^+ + Y - A^+ AYAA^+, Y \in C^{n\times m}. \quad (8.4.6)$$

注意到，$A^+\in\mathscr{M}(A^*, A^*)$，$Y-A^+AYAA^+\in\mathscr{N}(A, A)$. 若取内积$(A,B)=\mathrm{tr}A^*B$，则空间 $\mathscr{M}(A^*, A^*)$和 $\mathscr{N}(A,A)$是 $C^{m\times n}$ 的正交补空间．于是由(8.4.6)知

$$X \in \mathscr{M}(A^*, A^*)\Leftrightarrow Y - A^+AYAA^+ = 0\Leftrightarrow X = A^+.$$

引理证毕.

现在我们证明收敛性定理.

定理8.4.1 设 $A\in C^{m\times n}\neq 0$，取初始值 X_0满足

$$X_0 \in \mathscr{M}(A^*, A^*), \quad (8.4.7)$$

$$\rho(R_0) < 1, \quad (8.4.8)$$

则当$k\to\infty$时，由(8.4.5)定义的序列$X_k\to A^+$，对应的残差序列

$$\|R_{k+1}\| \leqslant \|R_0\|\|R_k\|, k = 0,1,2, \cdots \quad (8.4.9)$$

这里$\|\cdot\|$表示任一相容范数.

证明 因 $X_0\in\mathscr{M}(A^*, A^*)$，则 X_0满足(8.4.4)，于是

$$\begin{aligned} X_{k+1} &= X_k + X_0(I - AX_k) \\ &= X_k + X_0R_k \\ &= X_k + X_0(P_A - AX_k), \quad k = 0,1,2,\cdots \end{aligned}$$

对应的残差阵

$$\begin{aligned} R_{k+1} &= P_A - AX_{k+1} \\ &= P_A - AX_k - AX_0R_k \\ &= R_k - AX_0R_k. \end{aligned}$$

因为 $P_AR_k=R_k$，所以，

$$
\begin{aligned}
R_{k+1} &= P_A R_k - AX_0 R_k \\
&= R_0 R_k \\
&= R_0 (R_0 R_{k-1}) \\
&= \cdots \\
&= R_0^{k+2}
\end{aligned}
\tag{8.4.10}
$$

由范数的相容性，得

$$
\|R_{k+1}\| \leqslant \|R_0\| \|R_k\|.
$$

于是(8.4.9)得证.

由引理8.4.1和(8.4.8)、(8.4.10)知，当 $k\to\infty$时，残差阵

$$
R_k = P_A - AX_k \to 0. \tag{8.4.11}
$$

据此，我们可以证明 X_k 的收敛性. 事实上，利用(8.4.10)，

$$
\begin{aligned}
X_{k+1} &= X_k + X_0 R_k \\
&= X_k + X_0 R_0^{k+1} \\
&= X_{k-1} + X_0 R_0^k + X_0 R_0^{k+1} \\
&= \cdots \\
&= X_0 (I + R_0 + R_0^2 + \cdots + R_0^{k+1}).
\end{aligned}
\tag{8.4.12}
$$

因为 $\rho(R_0)<1$，依引理8.4.1，右端括号内矩阵级数是收敛的. 所以，X_{k+1}是收敛的. 记其极限为 X_∞，由(8.4.11)，立即得

$$
AX_\infty = P_A.
$$

但 $AX_\infty A = P_A A = A$，于是 $X_\infty\in A\{1\}$. 从 X_k 的定义及 X_0的假设知，对所有 k，$X_k\in \mathscr{M}(A^*, A^*)$，因而，$X_\infty\in \mathscr{M}(A^*, A^*)$，由引理8.4.2知，$X_\infty = A^+$. 定理证毕.

若 $A\in C^{m\times n}$，迭代公式(8.4.5)宜于用在 $m\leqslant n$ 的情形. 因为 R_k 和 $I_m - AX_k$ 都是 $m\times m$ 阵，如果 $m>n$，我们则从关系式

$$P_{A^*} = A^+A$$

出发，用 X_k 作为 A^+ 的近似，则对应的残差阵为

$$R_k = P_{A^*} - X_kA.$$

取初始阵 X_0 满足 $P_{A^*}X_0 = X_0$（例如 $X_0 \in \mathscr{M}(A^*, A^*)$），则相应的迭代序列应为

$$X_{k+1} = X_k + (I - X_kA)X_0.$$

如此，R_k 和 $(I_n - X_kA)$ 皆为 $n \times n$ 阵，定理8.4.1的结论仍成立.

利用引理8.4.1和(8.4.12)还可知

$$A^+ = \lim_{k\to\infty} X_{k+1} = X_0(I - R_0)^{-1}. \qquad (8.4.13)$$

初始值 X_0 的一个常用的选择是 $X_0 = \alpha A^*$，其中 α 为适当选择的实数. 因 $X_0 = \alpha A^*(AA^*)^- AA^*$，所以 $X_0 \in \mathscr{M}(A^*, A^*)$. 即对这样选择的 X_0，定理8.4.1中的条件(8.4.7)成立. 相应的迭代关系为

$$\begin{aligned} X_{k+1} &= X_k + X_0(I - AX_k) \\ &= \alpha A^* + (I_n - \alpha A^*A)X_k. \end{aligned} \qquad (8.4.14)$$

下面的定理刻画了这个迭代的收敛性质.

定理8.4.2 设 $A \in C_r^{m\times n}$，$\lambda_1(A^*A) \geqslant \cdots \geqslant \lambda_r(A^*A)$ 为 A^*A 的非零特征值，取初始值 $X_0 = \alpha A^*$.

(a) 假设

$$0 < \alpha < \frac{2}{\lambda_1(A^*A)}, \qquad (8.4.15)$$

则对(8.4.14)定义的序列 X_{k+1}，有

$$\lim_{k\to\infty} X_{k+1} = A^+.$$

(b) 残差矩阵 $R_k = P_A - AX_k$ 等于

$$R_k = (P_A - \alpha AA^*)^{k+1}, \quad k = 0,1,2,\cdots \qquad (8.4.16)$$

(c) 对任意的 k，谱范数$\|R_k\|_2$当

$$\alpha=\frac{2}{\lambda_1(A^*A)+\lambda_r(A^*A)} \tag{8.4.17}$$

时达到最小，其最小值为

$$\|R_k\|_2=\left[\frac{\lambda_1(A^*A)-\lambda_r(A^*A)}{\lambda_1(A^*A)+\lambda_r(A^*A)}\right]^{k+1},\quad k=0,1,2,\cdots \tag{8.4.18}$$

证明 (a) 因为 P_A 与 A^*A 可交换，所以它们可以用酉阵同时对角化(见定理2.3.5)。又因 $P_A\geqslant R_0=P_A-\alpha AA^*$ (这里 $A\geqslant B$ 表示 $A-B\geqslant 0$)，再结合 P_A 的特征值只能为1或0，可以证明 R_0的非零特征值只能是

$$1-\alpha\lambda_i(A^*A),\quad i=1,\cdots,r$$

中的若干个。因而

$$\rho(R_0)<1\Leftrightarrow|1-\alpha\lambda_i(A^*A)|<1$$
$$\Leftrightarrow(8.4.15)\text{ 成立}.$$

因此，当(8.4.15)成立时，定理8.4.1的条件(8.4.8)满足，再由 $X_0=\alpha A^*$，条件(8.4.7)也成立。故有 $X_k\to A^+$。

(b) 从(8.4.14)可以推出

$$X_k=\alpha\sum_{i=0}^{k}(I_n-\alpha A^*A)^iA^*, \tag{8.4.19}$$

经直接计算便可验证(8.4.16)。

(c) 因为 R_k 是 Hermite 阵，故

$$\|R_k\|_2=\rho(R_k)=\rho(R_0^{k+1})=\rho^{k+1}(R_0). \tag{8.4.20}$$

因此，$\|R_k\|_2$和 $\rho(R_0)$有相同的极小值点。对满足条件(8.4.15)的 α，$R_0=P_A-\alpha A^*A$ 的非零特征值为

$$1-\alpha\lambda_i(A^*A),\quad i=1,\cdots,r.$$

所以，α 要使 $\rho(R_0)=\max\{|1-\alpha\lambda_i(A^*A)|,i=1,\cdots,r\}$达到

最小的充要条件为

$$-[1-\alpha\lambda_1(A^*A)]=1-\alpha\lambda_r(A^*A).$$

这等价于(8.4.17). 注意到,对(8.4.17)选用的 α,

$$\rho(R_0)=|1-\alpha\lambda_r(A^*A)|.$$

于是

$$\|R_k\|_2=\rho^{k+1}(R_0)=|1-\alpha\lambda_r(A^*A)|^{k+1},$$

将(8.4.17)代入,即得到(8.4.18). 定理证毕.

这个定理表明,若取初始值

$$X_0=\frac{2}{\lambda_r(A^*A)+\lambda_1(A^*A)}A^*, \qquad (8.4.21)$$

应用迭代关系(8.4.14)所得到的序列 X_k 是收敛的,且 $X_k\to A^+$. 若选用 k 步迭代,则残差矩阵 R_k 由(8.4.16)给出,其谱范数为(8.4.18). 谱范数 $\|R_k\|_2$ 度量了 k 步迭代的精确程度. 若定义

$$K(A)=\frac{\lambda_1(A^*A)}{\lambda_r(A^*A)},$$

称为 A 的条件指数,则

$$\|R_k\|_2=\left(1-\frac{2}{K(A)+1}\right)^{k+1}.$$

可见, $\|R_k\|_2$ 是 $K(A)$ 的单调增函数. 因此 $K(A)$ 愈小,迭代收敛得愈快.

§8.5 其 他 方 法

本节讨论计算 A^+ 的极限方法、Greville 递推法以及奇异值分解法.

1. 极限方法

定理8.5.1 设 $A\in C^{m\times n}$,则

$$A^+=\lim_{\lambda\to 0^+} A^*(\lambda I_m+AA^*)^{-1} \qquad (8.5.1)$$

$$=\lim_{\lambda\to 0^+}(\lambda I_n+A^*A)^{-1}A^*. \qquad (8.5.2)$$

证明 设 A 有奇异值分解

$$A=P\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}Q^*,$$

其中 P 和 Q 分别为 $m\times m,n\times n$ 酉阵,$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_t),\sigma_i>0,i=1,\cdots,t$ 为 A 的奇异值,于是

$$AA^*=P\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}P^*,$$

$$\lambda I_m+AA^*=P\left(\lambda I_m+\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}\right)P^*,$$

$$A^*(\lambda I_m+AA^*)^{-1}=QD(\lambda)P^*,$$

其中 $D(\lambda)=\mathrm{diag}(d_1,\cdots,d_m)$,

$$d_i=\begin{cases}\dfrac{\sigma_i}{\lambda+\sigma_i^2}, & i=1,\cdots,t,\\ 0, & i=t+1,\cdots,m.\end{cases}$$

当 $\lambda\to 0^+$ 时,$d_i\to\dfrac{1}{\sigma_i}$,于是

$$A^*(\lambda I_m+AA^*)^{-1}\to Q\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{pmatrix}P^*=A^+.$$

同法可证(8.5.2).

2. Greville 递推法

设 $A=(a_1,\cdots,a_n)$ 为 $m\times n$ 阵,记 A_k 为 A 的前 k 列组成的 $m\times k$ 阵,将 A_k 分块为

$$A_k=(A_{k-1},a_k),\qquad k=2,\cdots,n.$$

约定 $A_1=a_1$,推论6.2.9表明

$$A_k^+ = \begin{pmatrix} A_{k-1}^+ - d_k b_k^* \\ b_k^* \end{pmatrix}, \quad k = 2, \cdots, n, \qquad (8.5.3)$$

其中

$$d_k = A_{k-1}^+ a_k.$$

$$c_k = a_k - A_{k-1} d_k,$$

$$b_k^* = \begin{cases} c_k^+, & 若\ c_k \neq 0. \\ (1 + d_k^* d_k)^{-1} d_k^* A_{k-1}^+, & 若\ c_k = 0. \end{cases} \qquad (8.5.4)$$

当我们计算 A^+ 时，先从 $A_1^+ = a_1^+$ 开始. 因为

$$A_1^- = a_1^+ = \frac{a_1^*}{a_1^* a_1},$$

有了 A_1^+，对 $k=2$，应用(8.5.4)和(8.5.3)计算出 A_2^+，然后再对 $k=3$用(8.5.4)和(8.5.3)计算出 A_3^+，这样递推下去，最后得到 $A_n^+ = A^+$.

这个方法是由 Greville(1960)提出的，他用这个方法计算 $A^+ y$(对任意 $y \in C^m$,)而不必计算 A^+.

3. 奇异值分解法

设 $A \in C_t^{m \times n}$，且有奇异值分解(见定理2.4.2)

$$A = P\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} Q^*,$$

这里 P 和 Q 分别为 $m \times m$，$n \times n$ 的酉阵，$\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_t)$，$\sigma_i > 0, i = 1, \cdots, t$. 我们知道

$$A^+ = Q\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} P^*, \qquad (8.5.5)$$

其中 Q 的列向量为 A^*A 的标准正交化特征向量，P 的列向量为 AA^* 的标准正交化特征向量，$\sigma_1^2, \cdots, \sigma_t^2$ 为 A^*A(或 AA^*)的正特征值. 若记

$$Q = (v_1, \cdots, v_n),$$

$$P = (u_1, \cdots, u_m),$$

则(8.5.5)可改写为

$$A^+ = \frac{1}{\sigma_1} v_1 u_1^* + \cdots + \frac{1}{\sigma_t} v_t u_t^*. \tag{8.5.6}$$

此式提供了通过计算 AA^* 和 A^*A 的标准正交化特征向量以及特征值而获得 A^+ 的一种方法.

第九章　概率统计中的应用

广义逆矩阵的重要应用领域之一是概率统计，特别在数理统计的线性模型、多元统计分析的参数估计方面，应用更为广泛. 正是这个原因，国外的一部分广义逆矩阵的专著，就出自统计学家之手. Pringle，R. M. 和 Rayner，A. A. (1971)的"Generalized Inverse Matrices with Applications to Statistics"，Rao，C. R. 和 Mitra，S. K. (1971)的"Generalized Inverse of Matrices and Its Applications"就是其中的两个例子.

本章扼要介绍广义逆在概率统计中的一些重要应用. 尽管应用相当广泛，但限于篇幅，这里只挑选若干方面作为示例，加以论述. 在§9.1，叙述协方差阵为半正定阵时多元正态分布(即奇异正态分布)的一些结果，包括条件分布，回归函数等以及一般多元分布当协方差阵奇异时，复相关系数和典则相关系数的表示. §9.2讨论正态变量二次型的服从 χ^2分布的充要条件. §9.3讨论线性模型的参数估计. §9.4介绍广义逆在判别函数表示中的应用. 通过这些讨论，读者将会看到，广义逆在数理统计的理论研究中，已经成为一个不可缺少的工具.

在本章的论述中，所有的向量和矩阵都是实的.

§9.1　奇异多元正态分布

多元正态分布是概率统计中最重要的统计分布. 关于它的定义，文献中有多种方法. 经典的方法是用密度函数，即若

$n\times 1$向量 x 具有密度函数

$$f(x)=\frac{1}{(2\pi)^{n/2}}\frac{1}{\det(\Sigma)^{1/2}}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad (9.1.1)$$

则称 x 服从均值为 μ,协方差阵为 Σ 的多元正态分布,记为 $x\sim N_n(\mu,\Sigma)$. 在这个定义中,用到了 Σ^{-1},因而要假设 Σ 为正定阵. 如果 $\Sigma\geqslant 0$,且 $\det\Sigma=0$,上面的定义就不能用了. 注意到正态随机向量 x 的分布,完全由它的一切可能的线性组合 $\alpha'x$ 的分布唯一确定,于是,我们可以用下面的方式来定义多元正态分布.

定义9.1.1 设 x 为 $n\times 1$随机向量,若对任意 $n\times 1$非随机向量 α,随机变量 $\alpha'x$ 的分布为一元正态,则称 x 的分布为多元正态分布.

多元正态分布的这种定义方法不涉及密度函数,因此称为"无密度方法"(density-free approach). 由这种定义,我们可以证明 x 的均值和协方差阵都存在,分别记为 $\mu=Ex$ 和 $\Sigma=\mathrm{Cov}(x)$. 因为它们完全确定了 x 的分布,所以可记为 $x\sim N_n(\mu,\Sigma)$,在不致引起混淆时,也简记为 $x\sim N(\mu,\Sigma)$. 若 $\det\Sigma=0$,即 Σ 为奇异阵时,x 的分布称为奇异正态分布. 本节我们讨论如何应用广义逆研究奇异正态分布的密度函数,条件分布以及一般情况下(即分布不必为正态时)的复相关系数和典则相关系数.

设 $x\sim N(\mu,\Sigma)$,$r(\Sigma)=r<n$,$Q=(Q_1\ \vdots\ Q_2)$为 Σ 的标准正交化特征向量组成的正交阵,Q_1为 $n\times r$ 矩阵且

$$Q'\Sigma Q=\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix},$$

这里 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_r)$,$\lambda_i>0$,$i=1,\cdots,r$,作变换 $y=Q'x$,则

$$y=\begin{pmatrix}y_{(1)}\\ y_{(2)}\end{pmatrix}=Q'x=\begin{pmatrix}Q_1'x\\ Q_2'x\end{pmatrix}\sim N_n\left(\begin{pmatrix}Q_1'\mu\\ Q_2'\mu\end{pmatrix},\begin{pmatrix}\Lambda & 0\\ 0 & 0\end{pmatrix}\right),$$

于是

$$y_{(1)} \sim N_r(Q_1'\mu, \Lambda), \Lambda > 0,$$

$$y_{(2)} \sim N_{n-r}(Q_2'\mu, 0). \tag{9.1.2}$$

对 $y_{(2)}$，我们有

$$Q_2'x = Q_2'\mu, \text{或 } Q_2'(x-\mu) = 0 \tag{9.1.3}$$

以概率为 1 成立.

因为 $\mathscr{M}(Q_2) = \mathscr{M}(Q_1)^{\perp} = \mathscr{M}(\Sigma)^{\perp}$，所以

$$(x-\mu) \in \mathscr{M}(Q_1) = \mathscr{M}(\Sigma) \tag{9.1.4}$$

以概率为 1 成立.

由于 $\Lambda > 0$，根据经典定义，$y_{(1)}$具有形如(9.1.1)的密度函数，只是将其中的 μ、Σ 和 n 分别换为 $Q_1'\mu$、Λ 和 r，即 $y_{(1)}$的密度函数为

$$g(y_{(1)}) = \frac{1}{(2\pi)^{r/2}} \frac{1}{\det(\Lambda)^{\frac{1}{2}}} e^{-\frac{1}{2}(y_{(1)}-Q_1'\mu)'\Lambda^{-1}(y_{(1)}-Q_1'\mu)}$$

$$= \frac{1}{(2\pi)^{r/2}} \frac{1}{\det(\Lambda)^{1/2}} e^{-\frac{1}{2}(x-\mu)'Q_1'\Lambda^{-1}Q_1(x-\mu)}.$$

因为 $\Sigma^{+} = Q_1'\Lambda^{-1}Q_1$，于是

$$g(y_{(1)}) = \frac{1}{(2\pi)^{r/2}} \frac{1}{\det(\Lambda)^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{+}(x-\mu)}.$$

注意到，在条件 $(x-\mu) \in \mathscr{M}(\Sigma)$ 下，$(x-\mu)'\Sigma^{-}(X-\mu)$与广义逆 Σ^{-} 选择无关，故

$$(x-\mu)'\Sigma^{+}(x-\mu) = (x-\mu)'\Sigma^{-}(x-\mu).$$

归纳起来，我们得到多元正态分布的密度.

定理9.1.1 设 $x \sim N_n(\mu, \Sigma)$，则

(a) 若 $\Sigma > 0$，则 x 具有密度(9.1.1).

(b) 若 $\det\Sigma = 0$，且 $r(\Sigma) = r$，则 $(x-\mu)$ 以概率为1落在子空间 $\mathscr{M}(\Sigma)$ 内（或者说在超平面 $Q_2'(x-\mu) = 0$ 上），且在此

子空间内，具有密度

$$g(\boldsymbol{x}) = (2\pi)^{-\frac{r}{2}} \frac{1}{|\boldsymbol{\Sigma}|_+^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^-(\boldsymbol{x}-\boldsymbol{\mu})},$$

其中$|\boldsymbol{\Sigma}|_+$表示$\boldsymbol{\Sigma}$的非零特征根的乘积.

推论9.1.1 设$\boldsymbol{x}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$，不论$\boldsymbol{\Sigma}>0$或$\boldsymbol{\Sigma}\geqslant 0$，$(\boldsymbol{x}-\boldsymbol{\mu})\in\mathscr{M}(\boldsymbol{\Sigma})$总是以概率1成立.

下面的定理是关于正态向量的子向量的条件分布.

定理9.1.2 若$\boldsymbol{x}\sim N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$，将$\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma}$分块为

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \qquad (9.1.5)$$

其中$\boldsymbol{x}_1,\boldsymbol{\mu}_1$为$k\times 1$向量，$\boldsymbol{\Sigma}_{11}$为$k\times k$矩阵. 记

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-\boldsymbol{\Sigma}_{21},$$

则

(a) $\boldsymbol{x}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-\boldsymbol{x}_2\sim N_k(\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{11.2})$，且与$\boldsymbol{x}_2$独立.

(b) 给定$\boldsymbol{x}_2$，$\boldsymbol{x}_1$的条件分布为

$$N_k(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11.2}).$$

证明 因为$\boldsymbol{\Sigma}\geqslant 0$，故存在矩阵$\boldsymbol{A}=(\boldsymbol{A}_1 \vdots \boldsymbol{A}_2)$，使得

$$\boldsymbol{\Sigma} = \boldsymbol{A}'\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}'_1\boldsymbol{A}_1 & \boldsymbol{A}'_1\boldsymbol{A}_2 \\ \boldsymbol{A}'_2\boldsymbol{A}_1 & \boldsymbol{A}'_2\boldsymbol{A}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

于是，$\mathscr{M}(\boldsymbol{\Sigma}_{21})\subset\mathscr{M}(\boldsymbol{A}'_2)=\mathscr{M}(\boldsymbol{A}'_2\boldsymbol{A}_2)=\mathscr{M}(\boldsymbol{\Sigma}_{22})$. 由此知，存在矩阵$\boldsymbol{B}$，使得$\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{22}\boldsymbol{B}$.

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^-\boldsymbol{\Sigma}_{22} = \boldsymbol{B}'\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{22}^-\boldsymbol{\Sigma}_{22} = \boldsymbol{B}'\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{12}. \qquad (9.1.6)$$

记

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{I}_k & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^- \\ \boldsymbol{0} & \boldsymbol{I}_{n-k} \end{pmatrix}$$

利用(9.1.6)，不难验证

$$C\Sigma C' = \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

因而

$$y = \begin{pmatrix} y_{(1)} \\ y_{(2)} \end{pmatrix} = Cx = \begin{pmatrix} x_1 - \Sigma_{12}\Sigma_{22}^{-}x_2 \\ x_2 \end{pmatrix}$$

$$\sim N\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right),$$

根据多元正态分布的基本性质(见王松桂,1987),(a)得证.(b)是(a)的直接推论. 定理证毕.

给定 x_2,x_1的条件分布的均值,即

$$E(x_1|x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2), \qquad (9.1.7)$$

称为 x_1对 x_2的回归函数. 其中 $\Sigma_{12}\Sigma_{22}^{-}$称为回归系数,(9.1.7)是线性回归. 注意到,由 $x-\mu\in\mathscr{M}(\Sigma)$以概率为1成立,可以推出 $x_2-\mu_2\in\mathscr{M}(\Sigma_{22})$以概率为1成立,所以(9.1.7)与广义逆 Σ_{22}^{-}的选择无关.

在下面的讨论中,我们只假设 $Ex=\mu$,$\mathrm{Cov}(x)=\Sigma$,并不要求它具有正态分布. 在(9.1.5)分块中,设 $k=1$,并记

$$\hat{x}_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2).$$

我们将随机变量 x_1与$(n-1)\times 1$随机向量 x_2的复相关系数定义为 x_1与 $\hat{x}_1$的简单相关系数,记之为 $R_{1.2,\cdots,n-1}$,即

$$R_{1.2,\cdots,n-1} = \frac{E(x_1-\mu_1)(\hat{x}_1 - E\hat{x}_1)}{\left[E(x_1-\mu_1))^2 E(\hat{x}_1 - E\hat{x}_1)^2\right]^{1/2}}. \qquad (9.1.8)$$

注意到,上式的分子

$$E(x_1-\mu_1)(\hat{x}_1 - E\hat{x}_1) = E(x_1-\mu_1)\Sigma_{12}\Sigma_{22}^{-}(x_2-\mu_2)$$

$$= \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21},$$

而

$$E(\hat{x}_1 - E\hat{x}_1)^2 = E[\Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2)(x_2 - \mu_2)'(\Sigma_{22}^{-})'\Sigma_{21}]$$
$$= \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21},$$
$$E(x_1 - \mu_1)^2 = \Sigma_{11},$$

将上面两式代入(9.1.8)得

$$R_{1.2,\cdots,n-1} = \left(\frac{\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}}{\Sigma_{11}}\right)^{1/2}, \tag{9.1.9}$$

这就是 x_1与其余 $n-1$个分量的复相关系数. 它的取值与广义逆 Σ_{22}^{-}的选择无关.

现在我们考虑两个随机向量 x_1和 x_2的相关程度的度量，这里 x_1和 x_2分别为 $k\times1$和$(n-k)\times1$随机向量，它们的协方差阵为 Σ. 和前面一样，我们把问题归结为随机变量的情况，即考虑线性组合 $\alpha' x_1$和 $\beta' x_2$的简单相关系数，这里 α 和 β 分别为 $k\times1$，$(n-k)\times1$向量.

不失一般性，可假设

$$\mathrm{Var}(\alpha' x_1) = 1, \mathrm{Car}(\beta' x_2) = 1. \tag{9.1.10}$$

于是，随机变量 $\alpha' x_1$与 $\beta' x_2$的相关系数为

$$\rho(\alpha' x_1, \beta' x_2) = \frac{\mathrm{Cov}(\alpha' x_1, \beta' x_2)}{\sqrt{\mathrm{Var}(\alpha' x_1)\mathrm{Var}(\beta' x_2)}} = \mathrm{Cov}(\alpha' x_1, \beta' x_2)$$
$$= \alpha' \Sigma_{12}\beta.$$

我们考虑在条件(9.1.10)下，求使 $\rho(\alpha' x_1, \beta' x_2)$达到最大值的 α 和 β.

因为(9.1.10)等价于 $\alpha' \Sigma_{11} \alpha = 1$，$\beta' \Sigma_{22} \beta = 1$，应用 Lagrange乘子法，构造辅助函数

$$F(\alpha, \beta, \lambda_1, \lambda_2)$$

$$= \alpha' \Sigma_{12}\beta - \frac{\lambda_1}{2}(\alpha' \Sigma_{11}\alpha - 1) - \frac{\lambda_2}{2}(\beta' \Sigma_{22}\beta - 1),$$

这里 λ_1和 λ_2为 Lagrange 乘子. 从$\frac{\partial F}{\partial \alpha}=0$和$\frac{\partial F}{\partial \beta}=0$,得

$$\Sigma_{12}\beta - \lambda_1\Sigma_{11}\alpha = 0 \tag{9.1.11}$$

$$-\lambda_2\Sigma_{22}\beta + \Sigma_{21}\alpha = 0 \tag{9.1.12}$$

这个方程组的一组解为

$$\lambda_1\alpha = \Sigma_{11}^{-}\Sigma_{12}\beta,$$

$$\lambda_2\beta = \Sigma_{22}^{-}\Sigma_{21}\alpha.$$

利用 $\Sigma_{11}\Sigma_{11}^{-}\Sigma_{12} = \Sigma_{12}$, $\Sigma_{22}\Sigma_{22}^{-}\Sigma_{21} = \Sigma_{21}$(参见(6.3.15)和(6.3.16)),可得

$$\lambda_1 = \lambda_1\alpha'\Sigma_{11}\alpha = \alpha'\Sigma_{11}\Sigma_{11}^{-}\Sigma_{12}\beta = \alpha'\Sigma_{12}\beta,$$

$$\lambda_2 = \lambda_2\beta'\Sigma_{22}\beta = \beta'\Sigma_{22}\Sigma_{22}^{-}\Sigma_{21}\alpha = \beta'\Sigma_{21}\alpha = \lambda_1.$$

因此,在(9.1.11)和(9.1.12)中,令 $\lambda_1 = \lambda_2 \overset{d}{=} \rho_1$,从(9.1.11)解出 $\lambda_1\alpha$ 代入(9.1.12),化简,我们得到

$$(\Sigma_{21}\Sigma_{11}^{-}\Sigma_{12} - \rho_1^2\Sigma_{22})\beta = 0.$$

注意,上式与所含广义逆的选择无关,故可将 Σ^{-} 取为 Σ^{+}. 可见,ρ_1^2是矩阵 $\Sigma_{21}\Sigma_{11}^{+}\Sigma_{12}$关于矩阵 Σ_{22}的相对特征值,而 β 为对应的特征向量. 但事实上,ρ_1^2也是 $\Sigma_{22}^{+}\Sigma_{21}\Sigma_{11}^{+}\Sigma_{12}$的特征值,这是因为若

$$(\Sigma_{22}^{+}\Sigma_{21}\Sigma_{11}^{+}\Sigma_{12} - \rho_1^2 I)\beta = 0, \tag{9.1.13}$$

用 Σ_{22}左乘上式,利用 $\Sigma_{22}\Sigma_{22}^{+}\Sigma_{21} = \Sigma_{21}$,得

$$(\Sigma_{21}\Sigma_{11}^{+}\Sigma_{12} - \rho_1^2\Sigma_{22})\beta = 0. \tag{9.1.14}$$

我们称 ρ_1^2的正平方根 ρ_1 为 x_1与 x_2的第一典则相关,而称 β 为 x_2-空间的典则向量,称 $\beta' x_2$为 x_2-空间的典则变量.

如果从(9.1.12)解出 $\lambda_2\beta$,代入(9.1.11),可得

$$(\Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} - \rho_1^2\Sigma_{11})\alpha = 0. \tag{9.1.15}$$

我们称该方程的解 α 为 x_1-空间的典则向量，称 $\alpha' x_1$ 为 x_1-空间的典则变量. 典则相关 ρ_1 为典则变量 $\alpha' x_1$ 和 $\beta' x_2$ 的简单相关系数. $\alpha' x_1$ 和 $\beta' x_2$ 是 x_1 和 x_2 所有可能线性组合中，具有最大相关系数的线性组合.

从上面的讨论我们可以看出，当分布的协方差阵为奇异阵时，广义逆就成为研究这种分布的一个不可缺少的工具.

§9.2 正态变量的二次型

设 $x \sim N_n(\mu, I)$，则随机变量 $y = x'x$ 的分布称为自由度为 n，非中心参数为 $\delta = \mu'\mu$ 的非中心 χ^2 分布，记之为 $y \sim \chi^2_{n,\delta}$. 若 $\delta = \mu'\mu = 0$，则称 y 的分布为中心 χ^2 分布，记为 $y \sim \chi^2_n$. 假设 A 为 $n \times n$ 实对称阵，b 为 $n \times 1$ 向量，c 为数，本节的目的是研究二次型 $x'Ax$ 和 $x'Ax + b'x + c$ 服从 χ^2 分布的条件.

定理9.2.1 设 $x \sim N_n(0, I)$，A 为实对称阵，则 $x'Ax \sim \chi^2_r$，当且仅当

$$A^2 = A, \qquad r(A) = r, \tag{9.2.1}$$

即 A 为秩为 r 的幂等阵.

证明 因为 A 为实对称阵，故存在 $n \times n$ 的正交阵 Q，使得 $Q'AQ = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，这里 $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值. 令 $y = Q'x$，则 $y \sim N_n(0, I)$，则

$$x'Ax = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \tag{9.2.2}$$

这里 $y_i \sim N_1(0,1)$，$i = 1, \cdots, n$，且相互独立. 于是，$y_i^2 \sim \chi^2_1$，$i = 1, \cdots, n$.

充分性 若(9.2.1)满足，则 A 的特征值只能是0和1，且1的个数为 r. 从(9.2.2)立得 $x'Ax \sim \chi^2_r$.

必要性 我们知道 χ^2_r 的矩母函数为

$$M_1(t) = (1-2t)^{-\frac{r}{2}}, |t| < \frac{1}{2} \qquad (9.2.3)$$

(关于矩母函数的定义及(9.2.3),参见方开泰等(1987)或 Mathai(1992)).而由(9.2.2)可推得 $x'Ax$ 的矩母函数为

$$M_2(t) = \prod_{j=1}^{n}(1-2\lambda_j t)^{-\frac{1}{2}}, |2\lambda_j t| < 1, j=1,\cdots,n, \qquad (9.2.4)$$

若 $x'Ax \sim \chi_r^2$,则应有 $M_1(t)=M_2(t)$,对满足 $|t|<\frac{1}{2}$, $|t|<\frac{1}{2|\lambda_j|}$, $j=1,2,\cdots,n$ 的一切 t 成立.将 $M_1(t)=M_2(t)$两边取对数,并将对数函数展开成 t 的无穷级数,令 $t,t^2,\cdots$的系数相等,得到一系列等式,其中之一为

$$r = \sum_{j=1}^{n}\lambda_j, \qquad (9.2.5)$$

将那一系列等式两两相减,我们得到一系列方程,其中之一为

$$\sum_{j=1}^{n} j\lambda_j^2(\lambda_j-1)^2 = 0, \qquad (9.2.6)$$

注意到 λ_j 是实数,因而 $\lambda_j^2(\lambda_j-1)^2=0$对一切 j 成立,故 $\lambda_j=0$ 或1,再从(9.2.5)知等于1的 λ_j 的个数为 r.这就证明了 A 是幂等阵,且 $r(A)=r$.定理证毕.

定理9.2.2 设 $x\sim N_n(\mu,I)$,A 为实对称阵,则

$$U = x'Ax + 2b'x + c \sim \chi_{r,\delta}^2$$

当且仅当

(a) $A^2=A$;

(b) $b\in\mathscr{M}(A)$,$c=b'b$,

其中 $r=r(A)$,$\delta=(b+\mu)'A(b+\mu)$.

证明 与定理9.2.1相同,存在正交阵 Q,使得 $Q'AQ=\text{diag}(\lambda_1,\cdots,\lambda_n)=\Lambda$,做变换 $y=Q'x$,则 $y\sim N_n(Q\mu,I)$,于是

$$U = y'\Lambda y + 2b'Qy + c.$$

用与定理9.2.1同样的方法，即比较 U 和 $\chi^2_{r,\delta}$ 的矩母函数可推出 $U \sim \chi^2_{r,\delta}$ 当且仅当

(a)′ $\lambda_i = 0$ 或 $1, i = 1, \cdots, r$;

(b)′ 若 $\lambda_i = 0$，有 $b_i = 0, c = b'b$，

这里 $r = \sum_i \lambda_i, \delta = \sum_i \lambda_i (b_i + \mu_i)^2$.

注意到(a)′⇔(a)，而结论"$\lambda_i = 0$ 必有 $b_i = 0$"$\Leftrightarrow Q'b \in \mathscr{M}(\Lambda) \Leftrightarrow b \in \mathscr{M}(A)$，同时

$$r = \sum_i \lambda_i = \operatorname{tr}\Lambda = \operatorname{tr}A = r(A),$$

$$\delta = (Q'b + Q'\mu)\Lambda(Q'b + Q'\mu) = (b' + \mu)'A(b + \mu).$$

定理证毕.

上面两个定理中，$\operatorname{Cov}(x) = I$，即 x 的分量是相互独立的. 下面我们讨论 $\operatorname{Cov}(x) = \Sigma$ 的情况，这里 $\Sigma \geqslant 0$，它可能是奇异的. 注意到，若 $\Sigma \geqslant 0, r = r(A)$，则存在矩阵 $B, r(B) = r$，使得 $\Sigma = BB'$. 依多元正态的性质(王松桂1974，p. 63)，x 可表为

$$x = \mu + By, \tag{9.2.7}$$

其中 $y \sim N_r(0, I)$. 因此，y 就是前面两个定理所讨论过的. 应用定理9.2.2，我们可以得到如下定理：

定理9.2.3 设 $x \sim N_n(\mu, \Sigma)$，则

$$U = x'Ax + 2b'x + c \sim \chi^2_{r,\delta}$$

当且仅当

(a) $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$;

(b) $\Sigma(A\mu + b) \in \mathscr{M}(\Sigma A \Sigma)$;

(c) $(A\mu + b)'\Sigma(A\mu + b) = \mu' A\mu + 2b'\mu + c$，

此时

$$r = \operatorname{tr}(A\Sigma),$$

$$\delta = (b + A\mu)'\Sigma A\Sigma(b + A\mu).$$

证明 将 x 的表示式(9.2.7)代入 U,得到

$$U = (\mu + By)'A(\mu + By) + 2b'(\mu + By) + c$$

$$= y'B'ABy + 2(A\mu + b)'By + \mu'A\mu + 2b'\mu + c.$$

根据定理9.2.2,$U \sim \chi^2_{r,\delta}$,当且仅当

(a)′ $B'AB$ 是幂等阵;

(b)′ $B'(A\mu+b) \in \mathscr{M}(B'AB)$;

(c)′ $(A\mu+b)'BB'(A\mu+b) = \mu'A\mu+2b'\mu+c$,

这里 B 的定义同(9.2.7.)

下面证明(a)′、(b)′、(c)′⇔(a)、(b)、(c).

从(a)′得

$$B'ABB'AB = B'AB. \tag{9.2.8}$$

用 B 和 B' 分别左乘和右乘上式,并利用 $\Sigma = BB'$ 得

$$\Sigma A\Sigma A\Sigma = \Sigma A\Sigma.$$

这就证明了(a)′⇒(a). 反过来,用 B' 和 B 分别左乘和右乘上式,并注意到 $B'B$ 是可逆的,我们得到(9.2.8). 这就证明了(a)⇒(a)′.

从(b)′知,存在 u,使得

$$B'(A\mu + b) = B'ABu.$$

因 B 为 $n\times r$,秩为 r 的矩阵,故上式等价于

$$BB'(A\mu + b) = BB'ABu.$$

因 $r(B) = r$,故存在 z,使得 $u = B'z$. 于是,上式变形为

$$\Sigma(A\mu + b) = \Sigma A\Sigma z,$$

此即(b),这就证明了(b)′⇒(b). 反过来,若(b)成立,则存在 t,使得 $\Sigma(A\mu + b) = \Sigma A\Sigma t$,将 $\Sigma = BB'$ 代入此式,并用 $(B'B)^{-1}B'$ 左乘之,便得到(b)′,这就证明了(b)′⇔(b),(c)⇔

(c)$'$是显然的. 定理证毕.

推论9.2.1 设 $x\sim N_n(0,\Sigma)$,则对广义逆 Σ^- 的任一选择,随机变量 $x'\Sigma^- x\sim\chi_r^2$,这里 $r=r(\Sigma)$.

根据上节的讨论,$x\in\mathscr{M}(\Sigma)$以概率1成立,于是 $x'\Sigma^- x$ 以概率为1与 Σ^- 的选择无关.

推论9.2.2 设 $x\sim N_n(0,\Sigma)$,则 $x'Ax\sim\chi_r^2$,当且仅当

$$\Sigma A\Sigma A\Sigma = \Sigma A\Sigma, \tag{9.2.9}$$

并且 $r=r(A)$.

若 $\Sigma>0$,则(9.2.9)可化简为

$$A\Sigma A = A, \tag{9.2.10}$$

即 Σ 为 A 的任一可逆的广义逆. 在两种情况下,均有 $r(A)=r$.

定理9.2.4 设 $x\sim N_n(\mu,\Sigma)$,若下列条件之一成立:

(a) $\mu\in\mathscr{M}(\Sigma)$;

(b) Σ^-是Σ的对称$\{1,2\}$-逆,

则 $x'\Sigma^- x$ 服从 $\chi_{r,\delta}^2$. 在任何一种情形,都有 $r=r(\Sigma)$, $\delta=\mu'\Sigma^-\mu$.

证明 对任一广义逆 Σ^-,定理9.2.3中的前两条对 $A=\Sigma^-$总是满足的. 而第三条,对现在的情形变为

$$(\Sigma^-\mu)'\Sigma(\Sigma^-\mu) = \mu'\Sigma^-\mu,$$

它等价于

$$\mu'(\Sigma^-)'\Sigma\Sigma^-\mu = \mu'\Sigma^-\mu.$$

而由本定理的条件(a)或(b)均可保证上式成立,定理证毕.

正态向量的二次型,出现在概率论与数理统计以及其他自然科学、工程技术的许多分支. 因此,关于它的研究,一直颇受人们的重视. 在这些研究中,广义逆往往是一个不可缺少的

工具. 本节概述了这方面的一些基本结果，对这一方向的近期发展想做更多了解的读者，可参阅 Khatri(1980).

§9.3 线 性 模 型

本节考虑一般线性模型

$$y = X\beta + \varepsilon, E(\varepsilon) = 0, \mathrm{Cov}(\varepsilon) = \sigma^2\Sigma \quad (9.3.1)$$

的参数估计和假设检验问题，这里 y 为 $n\times 1$观测向量，X 为 $n\times p$ 设计矩阵，$r(X)=r\leqslant p$，β 为 $p\times 1$回归系数向量，ε 为 $n\times 1$随机误差. 我们的讨论将表明，广义逆矩阵在模型(9.3.1)的统计推断中，起着十分重要的作用.

定义9.3.1 设 $c\in R^p$，若存在 y 的线性函数 $a'y$，使得 $E(a'y)=c'\beta$，则称 $c'\beta$ 是可估的. 若在 $c'\beta$ 的一切线性无偏估计中，$a'y$ 的方差最小，则称 $a'y$ 为 $c'\beta$ 的最佳线性无偏估计(The Best Linear Unbiased Estimator)，以下简记为 BLUE.

定理9.3.1 下列命题是等价的：

(a) $c'\beta$ 是可估的；

(b) $c\in \mathscr{M}(X')$；

(c) $c'X^-X=c'$.

证明 (a)⇔(b) 依定义，$c'\beta$ 可估 ⇔ $\exists\ a\in R^n$，使得 $E(a'y)=c'\beta$，对一切 $\beta\in R^p$ 成立

⇔存在 $a\in R^n$，$a'X\beta=c'\beta$，对一切 $\beta\in R^p$ 成立

⇔$\exists\ a\in R^n$，$X'a=c$， (9.3.2)

这就证明了(a)⇔(b).

(b)⇔(c) 首先，(c)⇒(b)是显然的. 反过来，若(9.3.2)成立，则

$$c'X^-X = a'XX^-X = a'X = c.$$

于是证明了(b)⇔(c). 定理证毕.

下面我们分三种情形来讨论模型(9.3.1)的参数估计.

情形1 我们首先考虑 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 的情形.

设 $\boldsymbol{a}'\boldsymbol{y}$ 为可估函数 $\boldsymbol{c}'\boldsymbol{\beta}$ 的任一线性无偏估计,即 $E(\boldsymbol{a}'\boldsymbol{y})=\boldsymbol{c}'\boldsymbol{\beta}$,这等价于

$$\boldsymbol{X}'\boldsymbol{a}=\boldsymbol{c}. \tag{9.3.3}$$

另一方面

$$\mathrm{Var}(\boldsymbol{a}'\boldsymbol{y})=\sigma^2\boldsymbol{a}'\boldsymbol{a},$$

于是,求可估函数 $\boldsymbol{c}'\boldsymbol{\beta}$ 的 BLUE,等价于在条件(9.3.3)下求 $\boldsymbol{a}$,使 $\boldsymbol{a}'\boldsymbol{y}$ 的方差达到最小. 这归结为求相容方程组(9.3.3)的最小范数解,这里范数是指欧氏范数 $\|\boldsymbol{a}\|=(\boldsymbol{a}'\boldsymbol{a})^{\frac{1}{2}}$. 根据定理5.3.3,这个最小范数解为

$$\boldsymbol{a}=(\boldsymbol{X}')^{(1,4)}\boldsymbol{c} \tag{9.3.4}$$

这里 $\boldsymbol{A}^{(1,4)}$ 表示 $\boldsymbol{A}$ 的{1,4}-逆,即 $\boldsymbol{A}$ 的最小范数广义逆,详细讨论见 § 5.3. 因而,可估函数 $\boldsymbol{c}'\boldsymbol{\beta}$ 的 BLUE 为

$$\boldsymbol{a}_0'\boldsymbol{y}=\boldsymbol{c}'((\boldsymbol{X}')^{(1,4)})'\boldsymbol{y}. \tag{9.3.5}$$

但容易验证关系:对任一矩阵 $\boldsymbol{A}$,

$$(\boldsymbol{A}')^{(1,4)}=(\boldsymbol{A}^{(1,3)})'. \tag{4.3.6}$$

因而

$$\boldsymbol{a}_0'\boldsymbol{y}=\boldsymbol{c}'\boldsymbol{X}^{(1,3)}\boldsymbol{y}, \tag{9.3.7}$$

这里 $\boldsymbol{A}^{(1,3)}$ 表示 $\boldsymbol{A}$ 的{1,3}-逆,即 $\boldsymbol{A}$ 的最小二乘广义逆,详细讨论见 § 5.2.

注意到,$\boldsymbol{c}'\boldsymbol{\beta}$ 为可估函数,$\boldsymbol{c}'\boldsymbol{X}^{(1,3)}$ 与最小二乘广义逆 $\boldsymbol{X}^{(1,3)}$ 的选择无关,可取为

$$\boldsymbol{X}^{(1,3)}=(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'.$$

记

$$\hat{\beta} = (X'X)^- X'y, \tag{9.3.8}$$

于是(9.3.7)就变为 $a'_0 y = c'\hat{\beta}$. $c'\hat{\beta}$ 也称为 $c'\beta$ 的最小二乘估计，此因用极小化$\|y-X\beta\|^2$的方法，可导致所谓正则方程

$$X'X\beta = X'y \tag{9.3.9}$$

这是一个相容方程组，其任一解可表为

$$\hat{\beta} = (X'X)^- X'y.$$

于是我们有如下定理：

定理9.3.2 在线性模型(9.3.1)中，设 $\Sigma = I$，则对任一可估函数 $c'\beta$，最小二乘估计 $c'\hat{\beta}$ 为其BLUE.

定理9.3.3 在线性模型(9.3.1)中，设 $\Sigma = I$，$c'_1\beta$ 和 $c'_2\beta$ 为两个可估函数，它们的BLUE分别为 $c'_1\hat{\beta}$ 和 $c'_2\hat{\beta}$)，则

$$\mathrm{Var}(c_1'\hat{\beta}) = \sigma^2 c'_1 (X'X)^- c_1,$$

$$\mathrm{Cov}(c'_1\hat{\beta}, c_2'\hat{\beta}) = \sigma^2 c_1' (X'X)^- c_2,$$

以上两式都与广义逆的选择无关.

证明 事实上，因为 $c_1'\beta$ 可估，故存在 α，使得 $c_1 = X'\alpha$，于是

$$\begin{aligned}\mathrm{Var}(c_1'\hat{\beta}) &= \mathrm{Var}(c_1'(X'X)^- X'y) \\ &= \sigma^2 c_1'(X'X)^- X'X((X'X)^-)'c_1 \\ &= \sigma^2 \alpha' X(X'X)^- X'X(X'X)^- X'\alpha \\ &= \sigma^2 c_1'(X'X)^- c_1,\end{aligned}$$

这里利用了

$$X((X'X)^-)'X' = X(X'X)^- X',$$

$$X(X'X)^- X'X = X.$$

类似地可证明第二个结论. 定理证毕.

情形2 若 $\Sigma \neq I$，但为已知正定阵.

此时求可估函数 $c'\beta$ 的 BLUE 的问题，归结为求方程 (9.3.3) 的最小范数解. 但此时的范数定义为 $\|a\| = (a'\Sigma a)^{1/2}$. 依定理5.3.6，最小范数解

$$a_0 = (X')^{(1,4)}_{(\Sigma)} c,$$

这里 $A^{(1,4)}_{(N)}$ 表示 A 的加权 {1, 4}-逆，详见 § 5.3. 于是 $c'\beta$ 的 BLUE 为

$$a_0' y = c'((X')^{(1,4)}_{(\Sigma)})' y. \tag{9.3.10}$$

利用类似于 (9.3.6) 的如下关系：

$$(A')^{(1,4)}_{(\Sigma)} = (A^{(1,3)}_{(\Sigma^{-1})})', \tag{9.3.11}$$

其中 $A^{(1,3)}_{(N)}$ 表示 A 的加权 {1, 3} -广义逆，详见 § 5.2. 我们得到

$$a'_0 y = c' X^{(1,3)}_{(\Sigma^{-1})} y. \tag{9.3.12}$$

类似于前面的讨论，因为 $c'\beta$ 可估，故 $c' X^{(1,3)}_{(\Sigma^{-1})}$ 与加权广义逆 $X^{(1,3)}_{(\Sigma^{-1})}$ 的选择无关，可取为

$$X^{(1,3)}_{(\Sigma^{-1})} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}.$$

记

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y,$$

则 (9.3.12) 表为

$$a_0' y = c'\tilde{\beta}. \tag{9.3.13}$$

我们也称 $c'\tilde{\beta}$ 为可估函数 $c'\beta$ 的广义最小二乘估计，此因它可以从 $\min(y - X\beta)'\Sigma^{-1}(y - X\beta)$ 获得. 上面的结果可表述为如下定理：

定理9.3.4 在线性模型 (9.3.1) 中，设 $\Sigma > 0$，则对任一可估函数 $c'\beta$，它的广义最小二乘估计 $c'\tilde{\beta}$ 是 $c'\beta$ 的 BLUE.

情形3 若 Σ 为已知半正定阵.

和情形2一样，可估函数 $c'\beta$ 的 BLUE 为

$$a_0'y = c'(X_{(\Sigma)}^{(1,4)})'y. \tag{9.3.14}$$

但与情形2不同的是，(9.3.11)不再成立. 要沿用前面的方法是困难的. 于是，对这种情形，我们换一个方式来讨论，即采用 Rao 最小二乘统一理论(参阅王松桂，1987或 Wang, S. G. 等 1994).

记 $T=\Sigma+XX'$，考虑 $(y-X\beta)'T^-(y-X\beta)$ 的极小化. 对此式关于 β 求导，并令其等于零，得到方程组

$$X'T^-X\beta = X'T^-y. \tag{9.3.15}$$

其解为

$$\tilde{\beta} = (X'T^-X)^-X'T^-y. \tag{9.3.16}$$

定理9.3.5 在线性模型(9.3.1)中，设 $\Sigma\geqslant 0$，记 $T=\Sigma+XX'$，则对任一可估函数 $c'\beta$，$c'\tilde{\beta}$ 为其 BLUE.

为了证明这个定理，我们先证明两个引理.

引理9.3.1 记 $T=\Sigma+XX'$ 其中 $\Sigma\geqslant 0$，则

(a) $\mathscr{M}(T)=\mathscr{M}(\Sigma \vdots X)$；

(b) $(y-X\beta)'T^-(y-X\beta)$，$X'T^-X$，$X'T^-y$ 都与广义逆的选择无关.

证明 (a) 因为 $\Sigma\geqslant 0$，Σ 可分解为 $\Sigma=QQ'$，故

$$T=\Sigma+XX'=(Q \vdots X)\begin{pmatrix} Q' \\ X' \end{pmatrix},$$

$$\mathscr{M}(T)=\mathscr{M}\left((Q \vdots X)\begin{pmatrix} Q' \\ X' \end{pmatrix}\right)=\mathscr{M}(Q \vdots X)=\mathscr{M}(\Sigma \vdots X).$$

(b) 因为对线性模型(9.3.1)，$y\in\mathscr{M}(\Sigma \vdots X)$ 以概率为1成立，利用事实：由 $\mathscr{M}(A)\subset\mathscr{M}(B)$ 可推知 $A'B^-A$ 与 B^- 的选择无关，便可证明(b).

引理9.3.2 对于线性模型(9.3.1),设 $a'y$ 为可估函数 $c'\beta$ 的一个无偏估计,则它为 $c'\beta$ 的 BLUE 当且仅当对任意满足 $E(b'y)=0$ 的 $b'y$,有 $\mathrm{Cov}(a'y,b'y)=0$.

证明 因为 $l'y$ 为 $c'\beta$ 的无偏估计当且仅当 l 是方程组

$$X'l = c \tag{9.3.17}$$

的解,并且 a 为它的一个特解. 故应用推论3.3.2知,(9.3.17)的通解可表为

$$\begin{aligned} l &= a + (I - (X')^- X')t \\ &\overset{\mathrm{d}}{=} a + b, \end{aligned}$$

这里 $b=(I-(X')^-X')t$,当 $t\in R^n$ 变化时,$b'y$ 构造了零的无偏估计的全体. 因为

$$\mathrm{Var}(l'y) = \mathrm{Var}(a'y) + \mathrm{Var}(b'y) + 2\mathrm{Cov}(a'y,b'y), \tag{9.3.18}$$

由此式,易见充分性得证.

对于必要性,我们采用反证法. 若存在一个 b_0,使得 $b_0'y$ 是零的无偏估计,但 $\mathrm{Cov}(a'y,b_0'y)=c\neq 0$,不妨设 $c<0$,则取 $b=tb_0$ 代替(9.3.18)中的 b,此时(9.3.18)变为 t 的二次三项式

$$\mathrm{Var}(l'y) = \mathrm{Var}(a'y) + t^2\mathrm{Var}(b_0'y) + ct.$$

因为 $c<0$,当 t 适当小时,这个二次三项式后两项的符号完全由一次项系数 c 的符号决定,即存在 t_0,使后两项为负值. 这就是说,对 $l_0=a+t_0b_0$

$$\mathrm{Var}(l_0'y) < \mathrm{Var}(a'y).$$

这与假设相矛盾,命题得证.

定理的证明 对任一可估函数 $c'\beta$,存在 $\alpha\in R^n$,使得 $c'=\alpha'X$. 因为 $\mathscr{M}(X)\subset\mathscr{M}(T)$,故 $X'T^-X$ 与广义逆 T^- 的选择无关,可取 T^- 为对称正定阵. 于是,

$$\mathscr{M}(X') = \mathscr{M}(X'T^- X),$$

因而

$$X(X'T^- X)^- X'T^- X = X,$$

故

$$E(c'\tilde{\beta}) = \alpha' X(X'T^- X)^- X'T^- X\beta = \alpha' X\beta = c'\beta,$$

这就证明了 $c'\tilde{\beta}$的无偏性.

下面应用引理3.9.2来证明 $c'\tilde{\beta}$的方差最小. 设 $b=(I-(X')^- X')t$,则 $b'y$ 为零的任一无偏估计. 因为

$$\begin{aligned}
\mathrm{Cov}(c'\tilde{\beta}, b'y) &= \sigma^2\alpha' X(X'T^- X)^- X'T^- \Sigma b \\
&= \sigma^2\alpha' X(X'T^- X)^- X'T^- (T - XX')(I - (X')^- X')t \\
&= \sigma^2\alpha' X(X'T^- X)^- X'T^- T(I - (X')^- X')t.
\end{aligned}$$

但 $X'T^- T=X'$, $X'(I-(X')^- X')=0$,故上式等于零. 定理证毕.

推论9.3.1 在线性模型(9.3.1)中,若 $\mathscr{M}(X)\subset\mathscr{M}(\Sigma)$,则对任一可估函数 $c'\beta$,$c'\tilde{\beta}$为其 BLUE,这里

$$\tilde{\beta} = (X'\Sigma^- X)^- X'\Sigma^- y,$$

其中 Σ^- 为 Σ 的任一{1}-逆.

对于 $\Sigma \geqslant 0$的情形,上述解法是由 C. R. Rao 于本世纪70年代提出的,他建议的 T 为 $T=\Sigma+XUX'$,这里 U 为任一对称半正定矩阵,满足 $r(T)=r(\Sigma \vdots X)$. 前面我们讨论的只是 $U=I$ 的特殊情形. 对于一般的 U,其证明完全相类似. Rao 把这个方法称为最小二乘统一理论(unified theory of least squares),其原因是定理9.3.5对一切线性模型都成立,即对一切可逆或不可逆的协方差阵 Σ 和一切列满秩或列降秩的设计阵 X 都成立.

对于 $\Sigma\geqslant 0$的情形,统计文献中还存在其他利用广义逆构

造可估函数的 BLUE 的方法，详见王松桂(1985).

广义逆在线性模型参数估计中的应用是多方面的，例如，在带线性约束的情形、方差分量模型、生长曲线模型等，广义逆都是不可缺少的工具. 本节只是对这一领域的应用做一概括介绍，对这方面感兴趣的读者可参阅王松桂(1987)和 Wang 等(1994).

§9.4 判 别 函 数

在多元统计分析中，有一种处理多元数据的常用方法，叫做判别分析. 它的目的是将一个个体"判入到已知的若干个个体类"中的一类. 例如，医生要根据一个人的若干项化验结果判断他是否患了某种疾病；地质学家要根据一块岩石标本的特征，判断它属于哪个地质年代；动植物检验人员往往要根据化验数据，判断一种样品属于有害群体类还是无害群体等等。从数学上抽象地说，假定我们有两个已知总体 π_1 和 π_2，我们希望建立一个判别准则，对给定的一个样本 x，依据这个准则，判断它是来自哪个总体，这就是判别分析要解决的问题.

本节的目的不是详细讨论判别分析问题，只是介绍广义逆在判别函数中的应用. 我们假设有两个多元正态总体，为简单计，假定它们有公共的协方差阵 $\boldsymbol{\Sigma}$,

$$\pi_1: N_p(\mu_1, \boldsymbol{\Sigma}),$$

$$\pi_2: N_p(\mu_2, \boldsymbol{\Sigma}),$$

这里 $\boldsymbol{\Sigma}$ 为 $p\times p$ 矩阵. 在一般多元统计分析教材中(如张尧庭和方开泰(1983)；王学仁和王松桂(1990)等)都讨论过 $\boldsymbol{\Sigma}>0$ 的情况. 本节假定 $r(\boldsymbol{\Sigma})=r<p$，即 $\boldsymbol{\Sigma}$ 为奇异阵，这在应用中常

常出现,例如,Krzanowski 等(1995)在试验光谱学、计算机辅助分子建模研究中所用的协方差阵就是奇异的.

假设 $\boldsymbol{Q}=(\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2)$ 为 $\boldsymbol{\Sigma}$ 的标准正交化特征向量组成的正交阵,$\boldsymbol{Q}_1$ 为 $p\times r$ 阵,$\boldsymbol{Q}_2$ 为 $p\times(p-r)$ 阵,且

$$\boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix},$$

这里 $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\cdots,\lambda_r)$,$\lambda_i>0$,$i=1,\cdots,r$,$r=r(\boldsymbol{\Sigma})$. 于是,根据 § 9.1知 π_i 以概率为1落在超平面

$$\boldsymbol{Q}_2'(\boldsymbol{x}-\boldsymbol{\mu}_i)=\mathbf{0}$$

上,且在此超平面上具有密度

$$f_i(\boldsymbol{x})=\frac{1}{(2\pi)^{r/2}}\frac{1}{|\boldsymbol{\Sigma}|_+^{1/2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^-(\boldsymbol{x}-\boldsymbol{\mu}_i)},i=1,2,$$

$$(9.4.1)$$

这里 $|\boldsymbol{\Sigma}|_+$ 表示 $\boldsymbol{\Sigma}$ 的非零特征值之积,即 $|\boldsymbol{\Sigma}|_+=\prod_{i=1}^{r}\lambda_i$.

我们分两种情况,建立它们的判别函数:

(a) 若 $\boldsymbol{Q}_2'\boldsymbol{\mu}_1=\boldsymbol{Q}_2'\boldsymbol{\mu}_2$.

即这两个总体在同一个超平面上,于是我们只需要从(9.4.1)出发,建立判别函数. 考虑 $\ln f_1(\boldsymbol{x})-\ln f_2(\boldsymbol{x})$,略去常数项以及与 $\boldsymbol{x}$ 无关的项,得到判别函数 $(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^-\boldsymbol{x}$,若记

$$d(\boldsymbol{x})=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^-\left(\boldsymbol{x}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\right),\qquad(9.4.2)$$

则对应的 Fisher 和 Bayes 判别准则为:若 $d(\boldsymbol{x})>0$,将样本 $\boldsymbol{x}$ 判入 π_1;若 $d(\boldsymbol{x})\leqslant 0$,则将样本 $\boldsymbol{x}$ 判入 π_2. $d(\boldsymbol{x})$ 的协方差阵为

$$\mathrm{Cov}d(\boldsymbol{x})=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\stackrel{\mathrm{d}}{=}\Delta,\quad(9.4.3)$$

这是奇异情形的 Mahalanobis 距离,它决定了判别效率. Δ 愈大,错判概率就愈小,判别效率就愈高.

(9.4.2)和(9.4.3)都包含了广义逆 $\boldsymbol{\Sigma}^-$,我们将证明

$d(x)$与Δ都跟Σ^-的选择无关.因为对现在这种情况,$Q_2'(\mu_1-\mu_2)=0$,故有

$$\mu_1-\mu_2\in\mathscr{M}(Q_1)=\mathscr{M}(\Sigma),\tag{9.4.4}$$

因而Δ与Σ^-选择无关.至于$d(x)$,我们只需证明,无论x来自π_1或是π_2都有

$$x-\frac{\mu_1+\mu_2}{2}\in\mathscr{M}(\Sigma).\tag{9.4.5}$$

事实上,若x来自π_1,

$$x-\frac{\mu_1+\mu_2}{2}=(x-\mu_1)+\frac{1}{2}(\mu_1-\mu_2),$$

因为$x-\mu_1\in\mathscr{M}(\Sigma)$,结合(9.4.4),便证明了(9.4.5).

类似地可以证明x来自π_2的情形.

(b) 若$Q_2'\mu_1\neq Q_2'\mu_2$.

此时这两个总体不在同一个超平面上,若x来自π_1,则

$$Q_2'x=Q_2'\mu_1$$

以概率1成立.若x来自π_2,则

$$Q_2'x=Q_2'\mu_2$$

以概率1成立.因此,$Q_2'x$就是很好的判别函数.

上面讨论的都是总体判别函数,即μ_i和Σ完全已知的情况.若μ_i,Σ未知,则需要训练样本,从训练样本获得μ_i,Σ的估计,就可得到样本判别函数.

第十章 其 他 应 用

在前面一些章节中，我们已经陆续讨论过广义逆矩阵在线性方程组、矩阵方程以及概率统计等方面的应用. 本章介绍广义逆矩阵在其他一些领域的应用，这包括线性规划、矩阵方程的整数解、非线性方程组以及网络理论等.

§10.1 区间线性规划

设 $\boldsymbol{A}=(a_{ij})\in \boldsymbol{R}^{m\times n}$，$\boldsymbol{a},\boldsymbol{b}\in \boldsymbol{R}^m$，$\boldsymbol{c},\boldsymbol{x}\in \boldsymbol{R}^n$. 线性规划问题

$$\begin{cases} a_i \leqslant \sum_{j=1}^{n} a_{ij}x_j \leqslant b_i, \quad i=1,\cdots,m, \\ \max\ \boldsymbol{c}'\boldsymbol{x} \end{cases}$$

称为区间线性规划(interval linear program)或带双边约束的线性规划(linear program with two-sided constraints). 若用 $\boldsymbol{a}\leqslant\boldsymbol{b}$ 表示 $a_i\leqslant b_i$，$i=1,\cdots,m$，则该线性规划问题可简记为

$$\begin{cases} \boldsymbol{a} \leqslant \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}, \\ \max\ \boldsymbol{c}'\boldsymbol{x}. \end{cases} \tag{10.1.1}$$

若 $\boldsymbol{x}$ 满足约束条件 $\boldsymbol{a}\leqslant\boldsymbol{A}\boldsymbol{x}\leqslant\boldsymbol{b}$，则称 $\boldsymbol{x}$ 为规划问题(10.1.1)的一个可行解. 记

$$S=\{\boldsymbol{x};\boldsymbol{a}\leqslant\boldsymbol{A}\boldsymbol{x}\leqslant\boldsymbol{b},\boldsymbol{x}\in\boldsymbol{R}^n\},$$

称 S 为规划问题(10.1.1)的可行解集. 若 S 非空，则称该规划问题是可行的. 如果

$$\max\{\boldsymbol{c}'\boldsymbol{x},\boldsymbol{x}\in S\}<\infty,$$

则称该规划问题是有界的. 如果存在 $\boldsymbol{x}_0\in S$，使得

$$c'x_0 = \max\{c'x, x \in S\},$$

则称 x_0 为区间线性规划问题(10.1.1)的最优解.

下面两个定理通过广义逆矩阵刻画了规划问题(10.1.1)的有界性及最优解集.

定理10.1.1 区间线性规划问题(10.1.1)有界，当且仅当 $c \in \mathcal{M}(A')$ 或等价地，$c'A^-A = c'$，对任一 A^- 成立.

证明 设 x 为任一可行解，$u \in \mathcal{N}(A)$，则容易验证 $x+u \in S$，因而 $S+\mathcal{N}(A)=S$，

$$\max\{c'x, x \in S\} = \max\{c'x, x \in S + \mathcal{N}(A)\}. \tag{10.1.2}$$

因为

$$P_{\mathcal{R}(A')} + P_{\mathcal{N}(A)} = I_n,$$

故(10.1.2)等于

$$\begin{aligned}
&\max\{(P_{\mathcal{R}(A')}c + P_{\mathcal{N}(A)}c)'x, x \in S + \mathcal{N}(A)\} \\
&= \max\{c'P_{\mathcal{R}(A')}x + c'P_{\mathcal{N}(A)}x, x \in S + \mathcal{N}(A)\} \\
&= \max\{c'P_{\mathcal{R}(A')}x, x \in S\} + \max\{c'P_{\mathcal{N}(A)}x, x \in \mathcal{N}(A)\} \\
&\stackrel{d}{=} l_1 + l_2,
\end{aligned}$$

这里

$$l_1 = \max\{c'P_{\mathcal{R}(A')}x, x \in S\},$$
$$l_2 = \max\{c'P_{\mathcal{N}(A)}x, x \in \mathcal{N}(A)\}.$$

利用

$$P_{\mathcal{R}(A')} = A^+A,$$

l_1 可改写为

$$l_1 = \max\{c'A^+Ax, x \in S\},$$

可见 l_1 是有限的. 另一方面，

$$l_2 = \max\{c'x, x \in \mathcal{N}(A)\},$$

因此，l_2有限$\Leftrightarrow c'x=0$，对一切 $x\in\mathscr{N}(A)\Leftrightarrow c\in\mathscr{N}(A)^{\perp}$. 不难证明

$$\mathscr{M}(I-A^{-}A)=\mathscr{N}(A),$$

对一切 A^{-}成立. 故 l_2有限$\Leftrightarrow c'(I-A^{-}A)=0$，对一切 A^{-}成立. 至于 $c\in\mathscr{M}(A')$与 $c'A^{-}A=c'$的等价性，证明是容易的. 定理证毕.

定理中条件"$c'A^{-}A=c'$，对一切 A^{-}成立"事实上等价于条件"$c'A^{-}A=c'$，对某一个特定的 A^{-}成立". 证明留给读者作练习.

定理10.1.2 假设区间线性规划问题(10.1.1)是可行有界的，$r(A)=m$. 则该问题的最优解集为

$$x=A^{-}u+(I-A^{-}A)t, \tag{10.1.3}$$

对任意 $t\in R^{n}$，其中 $u\in R^{m}$，且

$$u_i=\begin{cases}a_i, & 若(c'A^{-})_i<0,\\ b_i, & 若(c'A^{-})_i>0,\\ \lambda_i a_i+(1-\lambda_i)b_i, & 若(c'A^{-})_i=0,0\leqslant\lambda_i\leqslant 1,\end{cases} \tag{10.1.4}$$

其中$(c'A^{-})_i$ 表示行向量 $c'A^{-}$的第 i 个分量，A^{-}表示 A 的任意一个广义逆.

证明 因为对任意一个 A^{-}，有

$$\mathscr{M}(A^{-}\ \vdots\ I-A^{-}A)=R^{n}.$$

于是任一个向量 $x\in R^{n}$ 都可表为

$$x=A^{-}u+(I-A^{-}A)t,\qquad u\in R^{m},t\in R^{n}.$$

因为 $r(A)=m$，应用定理3.2.2有 $AA^{-}=I_m$，故

$$Ax=AA^{-}u=u. \tag{10.1.5}$$

根据该规划问题的有界性，利用上一定理，知 $c\in\mathscr{M}(A')$，于

是

$$c'x = c'(A^- u + (I - A^- A)t) = c'A^- u. \tag{10.1.6}$$

从(10.1.5)和(10.1.6),易知(10.1.3)和(10.1.4)给出了规划问题的最优解. 定理证毕.

本定理假设了 $r(A)=m$,即 A 是行满秩矩阵. Rao 和 Mitra(1971)讨论了 $r(A)<m$ 的情形.

上面我们以区间线性规划为例,说明了广义逆在这一领域的应用. 关于广义逆在一般规划问题中应用的更广泛结果,读者可参阅 Ben-Israel(1976)以及其后所引的文献.

§10.2 矩阵方程的整数解

一个矩阵 $A=(a_{ij})$ 若它的所有元素 a_{ij} 皆为整数,则称 A 为整数阵. 假设 A、B 和 D 皆为整数阵,考虑矩阵方程

$$AXB = D, \tag{10.2.1}$$

显然,若该矩阵方程相容,它的解未必是整数阵. 这里存在三种情况:或无整数解;或一部分解为整数解;或全部解都是整数解. 本节将讨论(10.2.1)何时有整数解,当存在整数解时,要找出它的全部或部分整数解. 线性方程组 $Ax=b$ 的整数解问题将作为(10.2.1)的特殊情形得到解决.

若一个整数矩阵 A 是可逆的,并且它的逆 A^{-1} 也是整数阵,则称 A 是单元阵(unit matrix)(参阅 Marcus 和 Minc, 1964, p. 42). 两个 $m\times n$ 整数阵 A 和 B 称为等价的,若存在 $m\times m$ 单元阵 P 和 $n\times n$ 单元阵 Q,使得

$$PAQ = B. \tag{10.2.2}$$

下面的引理证明了，任一整数阵（在上述意义下）一定等价于一整数对角阵.

引理10.2.1 设 A 为 $m\times n$ 整数阵，$r(A)=r$. 则 A 等价于 $m\times n$ 的整数对角阵 $S=(s_{ij})$，$r(S)=r$，且

(a) $s_{ii}\neq 0, i=1,\cdots,r$；

(b) $s_{ij}=0$，对所有其它 i,j；

(c) s_{ii} 可以整除 $s_{i+1,i+1}, i=1,\cdots,r-1$.

证明 首先找出 A 的所有元素的最大公因子，再对 A 施以如下两种初等行、列运算：

(1)交换两行或两列；

(2)从某一行(列)减去另一行(列)的整数倍，

把这个最大公因子移到(1,1)元位置，并使第一行和第一列的所有其它元素全变为0.

将所得到的矩阵记为 $B=(b_{ij})$. 显然，它等价于 A，且 b_{11} 能够除尽所有其它 b_{ij}. 令 $s_{11}=b_{11}$，则 B 有如下形式：

$$B=\begin{pmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B_1 & \\ 0 & & & \end{pmatrix},$$

这里 B_1 为 $(m-1)\times(n-1)$ 整数阵. 对 B_1 重复上面的运算，如此这样做下去，该运算重复 r 次，右下角 $(m-r)\times(n-r)$ 的子矩阵化为零阵.

在(10.2.2)中 $m\times m$ $(n\times n)$ 的矩阵 $P(Q)$ 等于所有用过的初等行(列)运算对应的初等矩阵的乘积，于是 P 和 Q 都是整数阵，它们的逆 P^{-1} 和 Q^{-1} 也都是整数阵，因而 P 和 Q 都是单元阵. 引理证毕.

引理中的矩阵 S 称为 A 的 Smith 正则形. 利用 Smith 正

则形，我们可以得到具有特殊整数性质的{1,2}-逆，这就是下面的引理.

引理10.2.2 设 A 为 $m\times n$ 整数阵，则存在 $A^{(1,2)}$ 记之为 B，使得 AB 与 BA 皆为整数阵.

证明 设 $PAQ=S$ 为 A 的 Smith 正则形. 记

$$B = QS^{+}P = Q\begin{pmatrix} \frac{1}{s_{11}} & & & & & \\ & \ddots & & & \mathbf{0} & \\ & & \frac{1}{s_{rr}} & & & \\ & \mathbf{0} & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} P, \qquad (10.2.3)$$

于是

$$PAQ = S = SS^{+}S = PAQS^{+}PAQ = PABAQ.$$

因为 P 和 Q 皆可逆，从上式立得 $A=ABA$. 类似地，可证 $BAB=B$. 这就证明了 $B\in A\{1,2\}$.

注意到

$$PAB = PAQS^{+}P = SS^{+}P,$$

故有

$$AB = P^{-1}SS^{+}P = P^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}P,$$

因为 P 和 P^{-1} 都是整数阵，于是 AB 也是整数阵. 类似地，从

$$BAQ = QS^{+}PP^{-1}SQ^{-1}Q = QS^{+}S = Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

可得

$$BA = Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1},$$

于是 BA 也是整数阵.引理证毕.

这个引理的证明是构造性的.即对任意 $m\times n$ 整数阵 A，它提供了构造一个 $A^{(1,2)}$ 的方法，使得 $AA^{(1,2)}$和 $A^{(1,2)}A$ 皆为整数阵.归纳起来，这个方法如下：

设 $r(A)=r$. 先用行和列的初等变换将 A 化为 Smith 正则形，即

$$\boldsymbol{PAQ} = \begin{bmatrix} s_{11} & & & & & \\ & \ddots & & & \boldsymbol{0} & \\ & & s_{rr} & & & \\ & \boldsymbol{0} & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \stackrel{\mathrm{d}}{=} \boldsymbol{S},$$

这里 P 和 Q 皆为单元阵. 定义

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{S}^{+}\boldsymbol{P},$$

此矩阵 B 即为所求的 $A^{(1,2)}$.

定理10.2.1 设 $A_i, i=1,2$和 D 为整数阵，且矩阵方程

$$\boldsymbol{A}_1\boldsymbol{X}\boldsymbol{A}_2 = \boldsymbol{D} \tag{10.2.4}$$

相容. 记 B_i 为引理10.2.2所找出的 A_i 的{1,2}-逆. 则

(a) 矩阵方程(10.2.4)有整数解当且仅当 B_1DB_2 为整数阵.

(b) 当(10.2.4)有整数解时，其整数解的通解可表为

$$\boldsymbol{X} = \boldsymbol{B}_1\boldsymbol{D}\boldsymbol{B}_2 + \boldsymbol{Y} - \boldsymbol{B}_1\boldsymbol{A}_1\boldsymbol{Y}\boldsymbol{A}_2\boldsymbol{B}_2, \tag{10.2.5}$$

这里 Y 为任一整数阵.

证明 (a) 若(10.2.4)相容，则存在矩阵 X_0，使得 $A_1X_0A_2=D$. 于是

$$\boldsymbol{A}_1(\boldsymbol{B}_1\boldsymbol{D}\boldsymbol{B}_2)\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{B}_1\boldsymbol{A}_1\boldsymbol{X}_0\boldsymbol{A}_2\boldsymbol{B}_2\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{X}_0\boldsymbol{A}_2 = \boldsymbol{D}.$$

这就证明了，若 B_1DB_2 为整数阵，则它是方程(10.2.4)的整数

解.

反过来，若(10.2.4)有整数解，设 X_0为一个这样的解，则

$$D = A_1X_0A_2 = A_1B_1A_1X_0A_2B_2A_2 = A_1B_1DB_2A_2,$$

于是 B_1DB_2为(10.2.4)的解.但

$$B_1DB_2 = B_1A_1X_0A_2B_2,$$

由引理10.2.2知，B_1A_1和 A_2B_2皆为整数阵，故 B_1DB_2也是整数阵.这就证明了(a).

(b) 当矩阵方程(10.2.4)有整数解时，根据(a)知，B_1DB_2必为整数解.显然，对任意的整数阵 Y，(10.2.5)定义的 X 也是整数阵.容易验证，它满足(10.2.4).反过来，若 X 是(10.2.4)的整数解，则它可表为

$$X = B_1DB_2 + X - B_1A_1XA_2B_2,$$

定理证毕.

把上面的定理用到线性方程组 $Ax=b$，我们得到如下推论.

推论10.2.1 设 A 为 $m\times n$ 整数阵，b 为 $m\times 1$整数向量，且方程组 $Ax=b$ 相容.记 B 为引理10.2.2找出的一个 $A^{(1,2)}$.则该方程组有整数解当且仅当 Bb 为整数向量.在有整数解的情形，其整数通解可表为

$$x = Bb + (I - BA)y,$$

其中 y 为任一 $n\times 1$整数向量.

§10.3 非线性方程组

本节介绍{2}-逆在非线性方程组迭代求解中的应用.

考虑非线性方程组

$$f_i(x_1,\cdots,x_n)=0,\quad i=1,\cdots,m. \qquad (10.3.1)$$

若记 $\boldsymbol{x}=(x_1,\cdots,x_n)'$，又记

$$\boldsymbol{f}(\boldsymbol{x})=\begin{pmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{pmatrix},$$

则(10.3.1)可以简单地表为

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{0}. \qquad (10.3.2)$$

设 $\boldsymbol{x}^*$ 为(10.3.2)的解，$\boldsymbol{x}_0$为其一个近似值，假定 $\boldsymbol{f}(\boldsymbol{x})$在 $\boldsymbol{x}_0$处可微，则

$$\boldsymbol{0}=\boldsymbol{f}(\boldsymbol{x}^*)\approx\boldsymbol{f}(\boldsymbol{x}_0)+D\boldsymbol{f}(\boldsymbol{x}_0)(\boldsymbol{x}^*-\boldsymbol{x}_0), \qquad (10.3.3)$$

其中

$$D\boldsymbol{f}(\boldsymbol{x})=\begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix},$$

这是一个 $m\times n$ 矩阵，称为 Jacobi 矩阵. 从(10.3.3)可以导出方程组

$$D\boldsymbol{f}(\boldsymbol{x}_0)(\boldsymbol{x}^*-\boldsymbol{x}_0)=-\boldsymbol{f}(\boldsymbol{x}_0), \qquad (10.3.4)$$

若 $m=n$ 且 $D\boldsymbol{f}(\boldsymbol{x}_0)$可逆，此方程组有唯一解

$$\boldsymbol{x}_1=\boldsymbol{x}_0-(D\boldsymbol{f}(\boldsymbol{x}_0))^{-1}\boldsymbol{f}(\boldsymbol{x}_0),$$

据此，我们可以构造如下迭代序列：

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-(D\boldsymbol{f}(\boldsymbol{x}_k))^{-1}\boldsymbol{f}(\boldsymbol{x}_k),\quad k=0,1,2,\cdots \qquad (10.3.5)$$

这就是解非线性方程组的 Newton 法迭代公式. 若 Jacobi 阵

$Df(x)$满足一定条件，可以证明 Newton 法至少是二阶收敛的. 关于这种情况的详细讨论见何旭初等(1980)，p. 214.

在(10.3.5)中，一个基本要求是，对每个 x_k，矩阵$Df(x_k)$是可逆阵. 这里产生一个问题，如果 $m\neq n$ 或 $m=n$，但 $Df(x_k)$不一定是可逆阵，此时(10.3.5)应该做怎样的修正？我们能不能使用 $Df(x_k)$的一个广义逆来代替它的普通逆？回答是肯定的. 这就是下面的定理.

设 $r>0$，$x_0\in C^n$，记

$$B(x_0,r)=\{x\in C^n:\|x-x_0\|<r\},$$

即 $B(x_0,r)$为以 x_0为中心，r 为半径的开球. 而$\overline{B(x_0,r)}$为此球的闭包，即以 x_0为中心、r 为半径的闭球.

定理10.3.1 对任意给定的 $\varepsilon>0$，$r>0$，$0<\delta<1$和 $A\in C^{m\times n}$，假设

(a) 对一切 $x,y\in B(x_0,r)$，有

$\|f(x)-f(y)-A(x-y)\|\leqslant\varepsilon\|x-y\|$；

(b) T 为一个 $A^{(2)}$，满足 $\varepsilon\|T\|=\delta<1$；

(c) $\|f(x_0)\|<\dfrac{(1-\delta)r}{\|T\|}$，

则序列

$$x_{k+1}=x_k-Tf(x_k) \tag{10.3.6}$$

是收敛的，且

$$\lim_{k\to\infty}x_k=x^*\in\overline{B(x_0,r)}, \tag{10.3.7}$$

$$Tf(x^*)=0, \tag{10.3.8}$$

其中$\|x\|$和$\|T\|$表示向量 x 和矩阵 T 的欧氏范数.

证明 不难看出，我们只需证明

$$x_k\in B(x_0,r) \tag{10.3.9}$$

$$\|x_{k+1}-x_k\|\leqslant\delta^k(1-\delta)r \tag{10.3.10}$$

对一切 $k=0,1,2,\cdots$成立.

对 k 施归纳法. 首先 $x_0\in B(x_0,r)$显然成立. 再由假设条件(c),有

$$\|x_1-x_0\|=\|Tf(x_0)\|$$
$$\leqslant\|T\|\|f(x_0)\|<(1-\delta)r,$$

这表明(10.3.9)和(10.3.10)对 $k=0$成立.

现在假设(10.3.9)和(10.3.10)对一切$0\leqslant j\leqslant k-1$成立. 由范数的三角不等式,可得

$$\|x_k-x_0\|\leqslant\sum_{j=0}^{k-1}\|x_{j+1}-x_j\|$$
$$\leqslant(1-\delta)r\sum_{j=0}^{k-1}\delta^j=(1-\delta^k)r<r,$$

即 $x_k\in B(x_0,r)$. 这就证明了(10.3.9)对 k 成立.

下面再证(10.3.10)对 k 成立. 利用(10.3.6),

$$x_{k+1}-x_k=-Tf(x_k)$$
$$=x_k-x_{k-1}-Tf(x_k)+Tf(x_{k-1}). \tag{10.3.11}$$

另一方面,从(10.3.6)知

$$x_k-x_{k-1}\in\mathscr{R}(T).$$

所以,存在 $u\in C^m$,使得 $x_k-x_{k-1}=Tu$. 于是由假设条件(b),有

$$TA(x_k-x_{k-1})=TATu=Tu=x_k-x_{k-1},$$

代入(10.3.11)得到

$$x_{k+1}-x_k=T[A(x_k-x_{k-1})-f(x_k)+f(x_{(k-1)})].$$

再利用范数的相容性以及假设(a),知

$$\|x_{k-1}-x_k\|\leqslant\|T\|\varepsilon\|x_k-x_{k-1}\|\leqslant\delta\|x_k-x_{k-1}\|$$
$$\leqslant\delta^{k+1}(1-\delta)r.$$

定理证毕.

对于迭代序列(10.3.6),本定理证明了它的极限 x^* 是 $Tf(x)=0$的解,它不必是 $f(x)=0$的解,仅当 $T\in C^{m\times m}$为列满秩阵(即 $r(T)=m$)时,x^*才是 $f(x)=0$的解.这是因为当 $r(T)=m$ 时T^*T 是可逆阵,用$(T^*T)^{-1}T^*$左乘(10.3.8),导致了 $f(x^*)=0$.我们知道,对任意一个 $A^{(2)}$,$r(A^{(2)})\leqslant r(A)\leqslant m$.可见,欲使 x^* 为 $f(x)=0$的解,一个必要条件是 $r(A)=m$,即假设条件(a)中的矩阵 A 必须是行满秩的.

我们称 $f(x)$在 $x_0\in C^n$ 处是可微的,若存在矩阵 A,使得

$$\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-A(x-x_0)\|}{\|x-x_0\|}=0,$$

这时 A 就是 Jacobi 阵 $Df(x_0)$.当 $f(x)$在 x_0可微时,对任意给定的 $\epsilon>0$,存在 $r>0$,当$\|x-x_0\|<r$ 时,

$$\|f(x)-f(x_0)-Df(x_0)(x-x_0)\|<\epsilon\|x-x_0\|. \tag{10.3.12}$$

若再要求 $Df(x)$在 $B(x_0,r)$上连续,则从(10.3.12)可推知定理10.3.1的假设(a)成立,于是我们证明了如下的推论.

推论10.3.1 对给定的 $x_0\in C^n$ 和 $\epsilon>0$,假设 $r>0$使得(10.3.12)成立.又设 $f(x)$在 $B(x_0,r)$上连续可微.对给定$0<\delta<1$,设 T 为 $Df(x_0)$的一个$\{2\}$-逆满足定理10.3.1的条件(b)和(c).则定理10.3.1的结论成立.

关于这一方向的更多结果,读者可参阅 Leach(1961),Altman(1955,1957),Ben-Israel(1965,1966a,1968),Rheinboldt(1968)以及 Ben-Israel 和 Greville(1974).

§10.4 网 络 理 论

本节的目的是介绍广义逆在电网络中的应用.

假设一个网络有 n 个终端，输入这 n 个终端的电流及对应的电压分别为 $i_1,\cdots,i_n$ 和 $v_1,\cdots,v_n$. 定义电流向量为 $\boldsymbol{i}=(i_1,\cdots,i_n)'$和电压向量为$\boldsymbol{v}=(v_1,\cdots,v_n)'$，它们之间的关系可线性地表为

$$\boldsymbol{i}=\boldsymbol{A}\boldsymbol{v}\ , \tag{10.4.1}$$

其中 $\boldsymbol{A}$ 为 $n\times n$ 不确定充入矩阵(indefinite admittance matrix)，它是实矩阵.

根据 Kirchoff 电流定律和电势的相对定律可以推出，$\boldsymbol{A}$ 满足关系

$$\boldsymbol{A}\mathbf{1}=\mathbf{0},\quad \mathbf{1}'\boldsymbol{A}=\mathbf{0}, \tag{10.4.2}$$

其中$\mathbf{1}=(1,\cdots,1)'$. 上式表明 $\boldsymbol{A}$ 的每一行元素之和等于零，同时，它的每一列元素之和也都等于零. 这样的矩阵称为双中心的(doubly centered). 从(10.4.2)知，$\boldsymbol{A}$ 就是一个奇异方阵. 于是，$\boldsymbol{A}^{-1}$不存在. 从(10.4.1)，我们可以形式地写出"逆"关系

$$\boldsymbol{v}=\boldsymbol{A}^{-}\boldsymbol{i}, \tag{10.4.3}$$

这里 $\boldsymbol{A}^{-}$是任一个{1}-逆. 问题是，是否存在一个 $\boldsymbol{A}^{-}$，使得(10.4.3)有一定物理意义?若记$\boldsymbol{v}=\boldsymbol{B}\boldsymbol{i}$，根据 Kirchoff 电流定律和电势相对定律，我们可以推知，$\boldsymbol{B}$ 也是双中心的，因此，在(10.4.3)中的 $\boldsymbol{A}^{-}$也应是双中心的. 最后，我们的问题归结为，在 $\boldsymbol{A}^{-}$中找一个双中心的阵，且使得(10.4.3)有一定物理意义.

引理10.4.1 设 $\boldsymbol{A}$ 为双中心的，且 $r(\boldsymbol{A})=n-1$，若 $\boldsymbol{A}\boldsymbol{X}=\mathbf{0}$，则存在向量 $\boldsymbol{u}$，使得

$$\boldsymbol{X}=\mathbf{1}\boldsymbol{u}'.$$

证明 从 $\boldsymbol{A}\boldsymbol{X}=\mathbf{0}$易知 $r(\boldsymbol{X})=1$. 于是，$\boldsymbol{X}$ 具有形式 $\boldsymbol{X}=\boldsymbol{a}\boldsymbol{u}'$. 再由 $\boldsymbol{A}\mathbf{1}=\mathbf{0}$，得 $\mathscr{M}(\boldsymbol{X})\subset\mathscr{M}(\mathbf{1})$，故得 $\boldsymbol{X}=\mathbf{1}\boldsymbol{u}'$. 证毕.

定理10.4.1 设 A 为双中心的 $n\times n$ 矩阵，$r(A)=n-1$，则 A 的唯一双中心广义逆是 A^+.

证明 设 X 为 A 的一个广义逆，即 $AXA=A$，等价地，

$$A(I-XA)=0.$$

利用引理10.4.1知存在 $n\times 1$ 向量 u，使得

$$I-XA=\mathbf{1}u'. \tag{10.4.4}$$

于是

$$\mathbf{1}'(I-XA)=\mathbf{1}'\mathbf{1}u'=nu'.$$

若 X 为双中心的，则 $\mathbf{1}'X=0$，故从上式得 $u=\mathbf{1}/n$. 代入(10.4.4)，得

$$XA=I-\frac{1}{n}\mathbf{1}\mathbf{1}',$$

这表明，XA 为实对称阵. 于是，我们证明了

$$X\in A\{1,4\}. \tag{10.4.5}$$

用类似方法可以证明 AX 也是实对称阵，因而，

$$X\in A\{1,3\}. \tag{10.4.6}$$

但 X 作为一个 A^-，必有 $r(X)\geqslant r(A)=n-1$，但从 $X\mathbf{1}=0$可以推出 $r(X)\leqslant n-1$. 这就证明了 $r(X)=n-1=r(A)$. 应用定理5.1.3，便有

$$X\in A\{1,2\}. \tag{10.4.7}$$

综合(10.4.5)、(10.4.6)和(10.4.7)，便证明了 $X=A^+$，定理证毕.

这个定理表明，在形为$v=A^-i$ 的"逆"关系中，$v=A^+i$ 是唯一满足 Kirchoff 电流定律和电势相对定律的逆关系.

下面的定理给出双中心矩阵的一些性质.

定理10.4.2 设 A 为 $n\times n$ 双中心阵，$r(A)=n-1$，则

(a) $(A+\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}\in A\{1\}$.

(b) $A^{+}=(A+\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}-\frac{1}{n}\mathbf{1}\mathbf{1}'=(A-\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}+\frac{1}{n}\mathbf{1}\mathbf{1}'$.

(c) $AA^{+}=A^{+}A=I-\frac{1}{n}\mathbf{1}\mathbf{1}'$.

(d) 若 A_1 和 A_2 为两个秩为 $n-1$ 的双中心阵，则

$$(A_1A_2)^{+}=A_2^{+}A_1^{+}.$$

证明 因为 $A\mathbf{1}=\mathbf{0}$，且 $r(A)=n-1$，于是 $A\pm\frac{1}{n}\mathbf{1}\mathbf{1}'$ 是可逆的. 因为 A 是双中心的，故

$$(A+\frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{1}=\mathbf{1}.$$

由上式，立得

$$(A+\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}\mathbf{1}=\mathbf{1}. \tag{10.4.8}$$

类似地，从

$$(A-\frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{1}=-\mathbf{1}$$

和

$$(A\pm\frac{1}{n}\mathbf{1}\mathbf{1}')(I-\frac{1}{n}\mathbf{1}\mathbf{1}')=A,$$

我们有

$$(A-\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}\mathbf{1}=-\mathbf{1} \tag{10.4.9}$$

$$(A\pm\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}A=I-\frac{1}{n}\mathbf{1}\mathbf{1}'. \tag{10.4.10}$$

利用这些关系，我们就可证明所要结论. 事实上

(a) 因为

$$(A+\frac{1}{n}\mathbf{1}\mathbf{1}')(A+\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}(A+\frac{1}{n}\mathbf{1}\mathbf{1}')=A+\frac{1}{n}\mathbf{1}\mathbf{1}',$$

将左边矩阵相乘，并利用(10.4.8)以及 $A\mathbf{1}=\mathbf{0}$和$\mathbf{1}'A=\mathbf{0}$，便得到

$$A(A+\frac{1}{n}\mathbf{1}\mathbf{1}')^{-1}A=A.$$

(b)和(c) 利用(10.4.8)、(10.4.9)和(10.4.10)容易证明所要结论.

利用(c)和定理4.3.1，不难证明(d). 定理证毕.

前面我们概括介绍了{1}-逆和 Moore-Penrose 广义逆在网络理论中的应用. 还有另外两种广义逆，即 Drazin 逆和 Bott-Duffin 逆，在网络理论中也很有用. 对这方面感兴趣的读者，可参阅 Ben-Israel 和 Greville(1974)，Ben-Israel(1976)以及后者所引的文献.

参考文献

［1］ 方开泰，许建伦．统计分布．北京：科学出版社，1987

［2］ 王松桂．线性模型参数估计的新进展．数学进展，1985，14：193—204

［3］ 王松桂．线性模型的理论及其应用．合肥：安徽教育出版社，1987

［4］ 王松桂，贾忠贞．矩阵论中不等式．合肥：安徽教育出版社，1994

［5］ 王学仁，王松桂．实用多元统计分析．上海：上海科技出版社，1990

［6］ 王国荣．上双对角阵 Moore-Penrose 逆及其并行算法．高校计算数学学报，1990，12：14—22

［7］ 王国荣．并行计算不相容线性方程组极小范数最小二乘解的 PCR 算法．高校计算数学学报（英文版），1993，2：1—10

［8］ 王国荣．矩阵与算子广义逆．北京：科学出版社，1994

［9］ 何旭初．广义逆矩阵的连续性问题——数值相关性的应用．高等学校计算数学学报，1979，1：168—172

［10］ 何旭初，苏煜城，包雪松．计算数学简明教程．北京：高等教育出版社，1980

［11］ 何旭初．广义逆矩阵的基本理论和计算方法．上海：上海科技出版社，1985

［12］ 孙继广．矩阵扰动分析．北京：科学出版社，1987

［13］ 许以超．代数学引论．上海：上海科技出版社，1965

［14］ 张尧庭，方开泰．多元统计分析引论．北京：科学出版社，1983

［15］ 陈希孺．数理统计引论．北京：科学出版社，1981

［16］ 倪国熙．常用的矩阵理论和方法．上海：上海科学技术出版社，1984

［17］ 缪建铭．秩 r 修正矩阵的 Moore-Penrose 逆．高校计算数学学报，1989，19：355—361.

［18］ Albert A. Regression and the Moore-Penrose Pseudoinverse.

New York: Academic Press, 1972

[19] Altman M. A generalization of Newton's method. Bull. Acad. Polon. Sci Ser. Sci. Math. Astronom. Phys., 1955, 3: 189—193

[20] Altman M. On a generalization of Newton's method. Bull. Acad. Polon. Sci, Ser. Sci. Math. Astronom. Phys., 1957, 5: 789—795

[21] Baksalary J K, Puntanen S. Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities. Aéquationes Math., 1991, 41: 103—110

[22] Baksalary J K, Nordstrom K, Styan G P H. Löwner ordering antitonicity of generalized inverses of Hermitian matrices. Linear Algebra and Applications, 1990, 127: 171—182

[23] Ben-Israel A. A modified Newton-Raphson method for the solution of systems of equations. Israel J. Math., 1965, 3: 95—98

[24] Ben-Israel A. A Newton-Raphson method for the solution of systems of equations. J. Math. Anal. Appl., 1966a, 15: 243—252

[25] Ben-Israel A. On error bounds for generalized inverses. SIAM J. Numer. Anal., 1966b, 3: 585—592

[26] Ben-Israel A. On applications of generalized inverses in nonlinear analysis. In Proceedings of the Symposium on Theory and Applications of Generalized Inverses of Matrices. eds. by Boullion T L and Odell P L. Lubbock: Texas Tech. Press, 1968, 183—202

[27] Ben-Israel A. Application of generalized inverses to programming games and networks. In Generalized Inverses and Applications, eds. by Nashed M Z. Academic Press, 1976, 495—524

[28] Ben-Israel A, Charnes A. Contributions to the theory of gener-

alized inverses. SIAM J. Appl. Math., 1963a, I :667—669

[29] Ben-Israel A, Charnes A. Generalized inverses and the Bott-Duffin network analysis. J. Math. Anal. Appl., 1963b, 7: 428—435

[30] Ben-Israel A, Charnes A. An explicit solution of a special class of linear programming problems. Operations Research, 1968, 16: 1167—1175

[31] Ben-Israel A, Greville T N E. Generalized Inverses: Theory and Applications. New York: John Wiley, 1974

[32] Bjerhammer A. Rectangular reciprocal matrices, with special reference to geodetic calculations. Bull. Géodésique, 1951a, 188—220

[33] Bjerhammer A. Applications of calculus of matrices to method of least squares with special reference to geodetic calculations. Kungl. Tekn. Hogsk. Handl, 1951b, No. 49, 36

[34] Bjerhammer A. Application of calculus of matrices to method of least squares with special reference to geodetic calculations. Kungl. Tekn. Hogsk. Handl. Stockholm, 1957, 49: 1—86

[35] Bjerhammer A. A generalized matrix algebra. Kungl. Tekn. Hogsk. Handl. Stockholm, 1968, 124: 1—32

[36] Bjerhammer A. Theory of Errors and Generalized Matrix Inverses. Amsterdam: Elsevier, 1973

[37] Bloomfield P, Watson G S. The inefficiency of least squares. Biometrika, 1975, 62: 121—128

[38] Bott R, R J Duffin. On the algebra of networks. Trans. Amer. Math. Soc., 1953, 74: 99—109

[39] Boullion T L, Odell P L. eds. *Proceedings of the Symposium on Theory and Application of Generalized Inverses of Matrices.* Mathematics Series, no. 4. Lubbock: Texas Tech. Press, 1968

[40] Boullion T L, Odell P L. Generalized Inverse Matrices. New York: John wiley, 1971

[41] Campbell S L. ed. Recent Applications of Generalized Inverses. Boston: Pitman, 1982

[42] Chipman J S. On least-squares with insufficient observations. J. Amer. Statist. Asoc., 1964, 59: 1078—1111.

[43] Chipman J S. Estimation and aggregation in econometrics: an application of the theory of generalized inverses. In Generalized Inverses and Applications (Nashed, M. Z., Ed). New York: Academic Press, 1976: 549—769

[44] Cline R E. Elements of the Theory of Generalized Inverses for Matrices. Newton: Education Development Center, 1979

[45] Cline R E, Pyle L D. The generalized inverse in linear programming an intersection projection method and the solution of a class of structured linear programming problems. SIAM J. Appl. Math., 1973, 24: 338—351

[46] Erdelyi I. The quasi-commuting inverses for a square matrix. Atti. Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. Ⅷ, 1967, 42: 626—633

[47] Erdelyi I. On the matrix equation $\mathbf{Ax} = \mathbf{Bx}$. J. Math. Anal. Appl., 1967, 17: 119—132

[48] Farebrother R W. Linear Least Squares Computations. New York: Marcel Dekker, 1988

[49] Goldberg J L. Matrix Theory with Applications. New York: McGraw-Hill, 1991

[50] Golub G H, Van Loan G F. Matrix Computations. Second Edition, Baltimore: Johns Hopkins University Press, 1989

[51] Getson A J, Hsuan F C. {2}-Inverses and Their Statistical Applications. New York: Springer-Verlag, 1980

[52] Graybill F A. Matrices with Applications in Statistics. Second

Edition, Belmont: Wadsworth, 1983

[53] Greville T N E. The pseudo inverse of a rectangular or singular matrix and its application to the solution of systems of linear equation. SIAM News Letter, 1957a, 5: 3—6

[54] Greville T N E. On smoothing a finite table: a matrix approach. SIAM J. Appl. Math., 1957b, 5: 137—154

[55] Greville T N E. Some applications of the pseudo inverse of a matrix. SIAM Review, 1960, 2: 15—22

[56] Horn R A, Johnson C R. Matrix Analysis. Cambridge: Cambridge University Press, 1985

[57] Khatri C G. Quadratic forms in normal variables. In Handbook of Statistics, Vol. 1, ed. by Krishnaiah P R. New York: North-Holland Publishing Company, 1980: 443—469

[58] Khatri C G, Rao C R. Some extensions of the Kantorovich inequality and statistical applications. J. Multi. Analysis, 1981, 11: 498—505

[59] Krzanowski W j *et al*. Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data. Applied Statistics, 1995, 44: 101—115

[60] Lavoie J L. A determinatal inequality involving the Moore-Penrose inverse. Linear Algebra and its Applications 1980, 31: 78—80

[61] Leach E B. A note on inverse function theorems. Proc. Amer. Math. Soc., 1961, 12: 694—697

[62] Liski E P, Pumtanen S, Wang S G (王松桂). Bounds for the trace of the difference of the covariance of the OLSE and BLUE. Linear Algebra and its Applications, 1992: 176: 121—130

[63] Liski E P, Wang S G (王松桂). On the 2-inverse and some ordering properties of nonnegative definite matrices. Acta.

Math. Appli. Sinica, 1996,12:22—27

[64] Liu S, Neudecker H. Kantorovich inequalities involving positive semidefinite matrices and efficiency comparisons in linear models. Preprint, 1995, to appear in Linear Algebra and its Applications

[65] Magnus J R, Neudecker H. Matrix Differential Calculus with Applications in Statistics and Econometrics. New York: John Wiley, 1991

[66] Marcus M, Minc H. A Survey of Matrix Theoty and Matrix Inequalities. Boston: Allyn & Bacon, 1964

[67] Marshall A W, Olkin I. Inequalities: Theory of Majorization and its Applications. New York: Academic Press, 1979

[68] Marshall A W, Olkin I. Matrix versions of the Cauchy and Kantorovich inequalities. Aéquations Mathematiae, 1990, 40: 89—93

[69] Mathai A M, Provost S B. Quadratic Forms in Random Variables. New York: Marcel Dekker Inc., 1992

[70] Milliken G A, Akdeniz F. A theorem on the difference of the generalized inverses of two nonnegative matrices. Commu. in Statist. Theory and Methods, 1977, 6:73—79

[71] Mitra S K, Rao C R. Simultaneous reduction of a pair of quadratic forms. Sankhyā., Ser. A, 1968a, 30:313—322

[72] Mitra S K, Rao C R. Some results in estimation and tests of linear hypothesis under the Gauss-Markov model. Sankhyā, Ser. A, 1968b, 30:281—290

[73] Mond B, Pecaric J E. Matrix versions of some means inequalities. Austral. Math. Soc. Gaz., 1993, 20: 117—120

[74] Moore E H. On the reciprocal of the general algebraic matrix (abstract). Bull. Amer. Math. Soc., 1920, 26:394—395

[75] Moore E H. General Analysis. Philadelphia: American Philo-

sophical Society, 1935

[76] Nashed M Z ed. Generalized Inverses and Applications: Proceedings of an Advanced Seminar. New York: Academic Press, 1976

[77] Noble B. A method for computing the generalized inverse of a matrix. SIAM J. Numer. Anal., 1966, 3:582—584

[78] Nordstrom K, Fellman J. Characterizations and dispersion matrix robustness of efficiently estimable parametric functionals in linear models with nuisances parameters. Linear Algebra and Its Applications, 1990, 127:341—361

[79] Nordstrom K. Some further aspects of the Löwner ordering antitonicity of the Moore-Penrose inverse. Commu. in Statist., Theory and Methods, 1989, 18:4471—4489

[80] Ouellette D V. Schur complements and statistics. Linear Algebra and its Applications, 1981, 36:187—295

[81] Pecaric J E, Puntanen S, Styan G P H. Some further matrix extensions of the Cauchy-Schwarz and Kantorovich inequalities. To appear in Linear Algebra and its Applications, 1996

[82] Penrose R. A generalized inverse for matrices, Proc. Cambridge hilos. Soc., 1955, 51:406—413

[83] Prasad K M. Generalized Inverses of Matrices over Rings. Ph Disertation. Bangalore: Indian Statistical Institute, 1992

[84] Pringle R M, Rayner A A. Generalized Inverse Matrices with Applications to Statistics. London: Charles Griffin, 1971

[85] Pyle L D. The generalized inverse in linear programming: basic structure. SIAM J. Appl. Math., 1972, 22:335—355

[86] Rao C R. Analysis of dispersion for multiply classified data with unequal numbers in cells. Sankhyā, 1955, 15:253—280

[87] Rao C R. A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. J. Roy.

Statist. Ser. B, 1962, 24:152—158

[88] Rao C R. Generalized inverse for matrices and its applications in mathematical statistics, Research Papers in Statistics, Festschrift for J. Neyman. New York: John Wiley,1966

[89] Rao C R. Calculus of generalized inverse of matrices, Part I: General theory. Sankhyā, Ser. A., 1967, 29:317—342

[90] Rao C R, Mitra S. K. Generalized Inverse of Matrices and its Applications. Wiley, New York: John Wiley, 1971

[91] Rheinboldt W C. A unified convergence theory for a class of iterative processes. SIAM J. Numer., 1968, 5:42—63

[92] Scroggs J E, Odell P L. An alternative definition of the pseudo inverse of a matrix. SIAM J. Appl. Math., 1966, 14:796—810

[93] Stewart G P W Ⅲ. On the continuity of the generalized inverse. SIAM J. Appl. Math., 1969, 17:33—45

[94] Stewart G W, Sun J G (孙继广). Matrix Perturbation Theory. Boston, Academic Press, 1990

[95] Styan G P H, Puntanen S. A Guide to Books on Matrices and on Inequalites, with Statistics & Other Applications. Report A 283, Dept. of Math. Science, Univ. of Tampere, Finland, 1993

[96] Uhlig F, Grone R eds. Current Trends in Matrix Theory: Proceedings of the Third Auburn Matrix Theory Conference. New York: Elsevier, 1987

[97] Wang Songgui (王松桂), Chow Shein-chung. Advanced Linear Models: Theory and Applications. New York: Marcel Dekker Inc., 1994

[98] Wang Songgui (王松桂), Shao Jun (邵军). Constrained Kantorovich inequalities and relative efficiency of least squares. J. Multi. Analysis, 1992, 42:284—298

[99] Wang Songgui（王松桂），Liski E P. Constraint Löwner ordering of matrices and extentions of the Cauchy-Schwarz inequality. Report A 304, Dept. of Math. Sciences, Univ. of Tampere, Finland, 1996

[100] Watkins D S. Fundamentals of Matrix Computations. New York: John Wiley, 1991

[101] Wedin P A. Perturbation theory for pseudoinverses. BIT, 1973, 13: 217—232

[102] Wu Chienfu. On some ordering properties of the generalized inverses of nonnegative definite matrices. Linear Algebra and its Applications, 1980, 32: 49—60

[103] Yang Zhenhai（杨振海），Liski E P. Goodness of fit tests based on binary data. Report A292, Dept. of Math. Sciences, Univ. of Tampere, Finland, 1995

索　引

三　画

四　画

五　画

六　画

七　画

八　画

九　画

十　画

十一　画

十二　画

十三 画以上

作者简介

王松桂 1965年毕业于中国科学技术大学数学系，现任北京工业大学教授，中国科学院应用数学所兼职研究员和博士生导师，中国科技大学兼职教授．长期从事数理统计、矩阵论等方面的科学研究．在《中国科学》、《科学通报》、《Linear Algebra and Its Applications》、《Annals of Statistics》等国内外刊物发表论文60余篇．出版的学术专著有《Advanced Linear Models》（英文专著，美国 Marcel Dekker 公司出版）、《线性模型的理论及应用》、《近代回归分析》、《矩阵论中的不等式》、《实用多元统计分析》等．曾先后应邀赴美、加拿大、日本、瑞典、瑞士、芬兰和波兰等国20余所大学讲学和合作研究．现是中国数学会、泛华国际统计协会、美国数学会会员，美国《数学评论》和德国《数学文摘》评论员．曾获中国科学院重大科技成果二等奖和北京市科技进步二等奖．

杨振海 1964年毕业于中国科学技术大学数学系，现任北京工业大学教授，中国数学会理事，中国现场统计研究会副理事长兼秘书长．长期从事概率论与数理统计、可靠性数学等领域的研究工作．已在国内外重要学术刊物发表论文30余篇．出版的学术专著有《数理统计基础》和《拟合优度检验》．曾先后应邀访问美、日本、芬兰和香港等国家或地区，进行合作研究或讲学．曾获全军科技进步一等奖、北京市学术奖和国家科技进步一等奖．1991年被国家教委和人事部授予"有突出贡献的留学归国人员"称号．